

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 7, 2004, 16:47:24 ; Search time 85.0565 Seconds
(without alignments)
833.792 Million cell updates/sec

Title: US-10-009-919A-1

Perfect score: 1271

Sequence: 1 MAKHKVRADELFLVQLGLAES.....PSVIKPGKNGQEVLIYLKXR 251

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A_Geneseq_29Jan04.*

1: geneseqp1980s.*

2: geneseqp1990s.*

3: geneseqp2000s.*

4: geneseqp2001s.*

5: geneseqp2002s.*

6: geneseqp2003as.*

7: geneseqp2003bs.*

8: geneseqp2004s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1271	100.0	251	4	AAB48003 L. intrac
2	446.5	35.1	270	2	AAW5094 Streptoco
3	446.5	35.1	270	5	ABP54588 S. pneumo
4	446.5	35.1	270	7	ADC45145 S. pneumo
5	431	33.9	274	5	ABB48304 Listeria
6	419.5	33.0	280	5	ABP65873 Bifidobac
7	416	32.7	275	5	ABP25556 Streptoco
8	406	31.9	275	5	ABP25555 Streptoco
9	383.5	30.2	240	2	AAR40752 Sequence
10	383.5	30.2	240	2	AAW4272 Sequence
11	381.5	30.0	240	2	AAR40751 Sequence
12	374.5	29.5	261	3	AAG23017 Arabidops
13	374.5	29.5	263	3	AAG51202 Arabidops
14	374.5	29.5	301	3	AAG23016 Arabidops
15	374.5	29.5	303	3	AAG51201 Arabidops
16	374.5	29.5	333	3	AAG23015 Arabidops
17	374.5	29.5	335	3	AAG51200 Arabidops
18	365	28.7	227	7	ADC97152 E. faeciu
19	363	28.6	243	5	ABP54174 Lactococc
20	358	28.2	268	6	ADB09362 Allostoco
21	358	28.2	271	6	ADB09364 Allostoco
22	347.5	27.3	273	4	AAG91307 C glutami
23	338.5	26.6	277	4	AAU45822 Propionib
24	338.5	26.6	277	6	ABM42341 Propionib
25	316	24.9	198	6	ABU01632 S. pneumo

ALIGNMENTS

RESULT 1

AAB48003

ID AAB48003 standard; protein; 251 AA.

XX

AC AAB48003;

XX

AC

DT 19-MAR-2001 (first entry)

XX

XX

DE L. intracellularis hemolysin polypeptide.

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

```
CC microorganisms, in hybridization or amplification assays
XX
SQ Sequence 251 AA;
  Query Match      100.0%; Score 1271; DB 4; Length 251;
  Best Local Similarity 100.0%; Pred. No. 1.4e-133;
  Matches 251; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MAKHKVRADELVFLQGLAESREQAQKRLIMAGKVTILTNNSSTIPLRLKPGHKYPLESICS 60
Db |||||
QY 61 LIGVERFVSRGAYKLLTALDFPKIDVKSCICLDAGASTGGFTDCLLQHGASKVYVIDVGK 120
Db |||||
QY 121 GOLHEKLYTNEOVINIEGVNLTASKDLIPEEVDILTIDVSFISLTLLPSCIRWLKASG 180
Db |||||
QY 181 IIIALIKQFELYPDKI-KKGVVKETSLOYEAVEKIIHFCQSELGLIFIGVWPSVIKGPKG 240
Db |||||
QY 241 NOEYLYLYLKKR 251
Db |||||

RESULT 2
AAW55094
ID AAW55094 standard; protein; 270 AA.
XX
AC AAW55094;
XX
DT 02-OCT-1998 (first entry)
XX
DE Streptococcus pneumoniae SP0041 protein.
XX
KW Streptococcus pneumoniae; antigen; vaccine; infection; diagnosis;
KW detection; pneumonia; otitis media; meningitis.
XX
OS Streptococcus pneumoniae.
XX
PW WO9818930-A2.
XX
PD 07-MAY-1998.
XX
PF 30-OCT-1997; 97WO-US019422.
XX
PR 31-OCT-1996; 96US-0029960P.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Kunsch CA, Choi GH, Johnson LS, Hromockyj A;
XX
DR WPI; 1998-272224/24.
DR N-PSDB; AAV27355.
XX
PT Nucleic acid encoding antigenic peptide(s) from Streptococcus pneumoniae
PT - or their epitope-containing fragments, useful in protective or
PT therapeutic vaccines, and for diagnosis.
XX
PS Claim 11; Page 61; 118pp; English.
XX
CC The present sequence represents a protein from Streptococcus pneumoniae.
CC The nucleic acid sequence encoding the Streptococcus pneumoniae protein
CC can be useful in vaccines for inducing protective antibodies against
CC Streptococcus pneumoniae, for treatment or prevention of infection e.g.
CC pneumonia, otitis media or meningitis. Probes based on the nucleic acid
CC are used to detect Streptococcus infection (by usual hybridisation or
CC amplification methods), also for isolating Streptococcus genes or their
CC allelic variants. The protein can be used similarly to detect specific
CC
```

```
CC antibodies in standard immunoassays, especially for diagnosing or
CC monitoring infections. Antibodies which bind the protein are used to
CC detect corresponding antigens, to purify the protein and for passive
CC immunisation (optionally coupled to a toxin). Vaccines are administered,
CC e.g. by injection, orally or through the skin, typically at 0.01-1000
CC (especially 10-300) mu g/ml per dose
XX
SQ Sequence 270 AA;
  Query Match      35.1%; Score 446.5; DB 2; Length 270;
  Best Local Similarity 42.7%; Pred. No. 4.3e-41;
  Matches 106; Conservative 41; Mismatches 94; Indels 7; Gaps 4;
QY 5 KVRADLVFLQGLAESREQAQKRLIMAGKVTILTNNSSTIPLRLKPGHKYPLESICSLIGV 64
Db |||||
QY 2 KERVDLVAYKQGLFETREQAKRGVWAGLVAVLNCE---RFDKPEKIPDDTELKKE 57
Db |||||
QY 65 E-RFVSRGAYKLLTALDFPKIDVKSCICLDAGASTGGFTDCLLQHGASKVYVIDVGKGL 123
Db |||||
QY 58 KLVYVSRLGLKLEKALQVDFLSDVGATTIDIGASTGGFTDVMQLQNSAKLVFAVDVGTQL 117
Db |||||
QY 124 HEKLYTNEQVINIEGVNLTASKDLIPEEVDILTIDVSFISLTLLPSCIRWLKASGIII 183
Db |||||
QY 118 AWKLEQDPRVVSMEQFNERYAEKTDFEQPSFASIDVSFISLTLLPALHRLVLAQGVV 177
QY 184 ALIKPQFELYPDKI-KKGVVKETSLOYEAVEKIIHFCQSELGLIFIGVWPSVIKGPKNQ 242
Db |||||
QY 178 ALVKPQFAGREGIQKNGIIRDAKHQNVLESVTAMA-VEVGFSLGLDFSPIQGHGNI 236
QY 243 EYLYLYLKK 250
Db |||||
Db 237 EFLAYLKK 244

RESULT 3
ABP54588
ID ABP54588 standard; protein; 270 AA.
XX
AC ABP54588;
XX
DT 04-SEP-2002 (first entry)
XX
DE S. pneumoniae SP041 protein sequence SEQ ID NO:64.
XX
KW Streptococcus pneumoniae; epitope; vaccine; antigenic protein;
KW antibacterial; Streptococcal infection; detection.
XX
OS Streptococcus pneumoniae.
XX
PN US2002061545-A1.
XX
PD 23-MAY-2002.
XX
PF 22-JAN-2001; 2001US-00765272.
XX
PR 30-OCT-1997; 97US-00961083.
XX
PA (CHOI/) CHOI G H.
PA (KUNS/) KUNSCH C A.
PA (BARA/) BARASH S C.
PA (DILL/) DILLON P J.
PA (DOUG/) DOUGHERTY B.
PA (FANN/) FANNON M R.
PA (ROSE/) ROSEN C A.
XX
PI Choi GH, Kunsch CA, Barash SC, Dillon PJ, Dougherty B, Fannon MR;
PI Rosen CA;
XX
DR WPI; 2002-479261/51.
DR N-PSDB; ABQ84823.
XX
PT New Streptococcus pneumoniae antigens, useful for detecting Streptococcus
PT and for preventing or attenuating disease caused by Streptococcus
```

```
PT infection.
XX Claim 11; Page 28; 70pp; English.
PS
XX
CC ABQ84792 to ABQ84904 represents nucleic acids which encode the
CC Streptococcus pneumoniae antigens given in ABP54557 to ABP54669. The S.
CC pneumoniae antigens have antibacterial activity and can be used in
CC vaccines. The S. pneumoniae antigens can also be used to prevent or
CC attenuate a Streptococcal infection in an animal. The polynucleotides
CC encoding the S. pneumoniae antigens can be used to detect Streptococcus
CC nucleic acids. ABQ84905 to ABQ85130 represent primers used in the cloning
CC of S. pneumoniae ORFs (open reading frames) which are used in an example
CC from the present invention
XX
SQ Sequence 270 AA;
Query Match 35.1%; Score 446.5; DB 5; Length 270;
Best Local Similarity 42.7%; Pred. No. 4.3e-41;
Matches 106; Conservative 41; Mismatches 94; Indels 7; Gaps 4;
QY 5 KYRADLVFLQGLAESREQAQRLIMAGKVTLTNNSTTIPRLLEKPGHKYPLESICSLIGV 64
DB 2 KERVDVLAYKQGLFETREQAKRGWAGLVVAVLNGE----RFDKPGEKIPDDTELKKG 57
QY 65 E-RFVSRGAYKLLTALDFFKIDVKSCICLDAGASTGGFTDCLLQHGASKYVYADVKGQL 123
DB 58 KLVYVSRGGKLEKALQVFDLSVDGATTIDIGASTGGFTDVM-LQNSAKLVFADVVGINQL 117
QY 124 HEKLYTNEQVINTEGVNLTASKDLIPEEVDILTIDVSFISLTILPSCIRWLKASGIII 183
DB 118 AWKLRQDPVRSMEQNFRYAEKTDPEQPSFASIDVSFISLSLILPALHRVLADQGVV 177
QY 184 ALIKPQFELYPDKI-KGVVVKETSLQYEAVEKIIHFQCSQELGLIFIGWVPSVIKPKGNQ 242
DB 178 ALVKPQFEAGREQIGKNGIIRDAKVHQNVLSTAMA-VEVGFSVLGLDFSPIQGGHNI 236
QY 243 EYLIYLKK 250
DB 237 EFLAYLKK 244
RESULT 4
ADQ45145
ID ADC45145 standard; protein; 270 AA.
XX
AC ADC45145;
XX
DT 18-DEC-2003 (first entry)
XX
DE S. pneumoniae antigenic protein SP041.
XX
KW Antigen; bacterial infection; vaccine; pneumonia; antibacterial.
XX
OS Streptococcus pneumoniae.
XX
XX
XX US6573082-B1.
XX
XX 31-OCT-1996; 96US-0029960P.
XX
XX 30-OCT-1997; 97US-00961083.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Choi GH, Kunsch CA, Barash SC, Dillon PJ, Dougherty B, Fannon MR;
XX Rosen CA;
XX
XX WPI; 2003-764574/72.
XX
XX N-PSDB; ADC45144.
XX
XX Novel polynucleotide encoding Streptococcus pneumoniae polypeptides
PT
```

```
PT useful for producing vaccines for prevention or attenuation of infection
XX by Streptococcus pneumoniae.
XX
PS Example 1; SEQ ID NO 64; 58pp; English.
XX
CC The invention relates to an isolated polynucleotide consisting of a
CC Streptococcus pneumoniae nucleic acid (appearing as ADC45122 and encoding
CC SP028) one of 113 disclosed nucleic acids encoding 113 S. pneumoniae
CC antigens. Also included are making a recombinant vector by inserting the
CC nucleic acid into a vector, an isolated polynucleotide consisting of at
CC least 50 or 100 contiguous nucleotides of the SP028 nucleic acid, and a
CC recombinant host cell comprising the SP028 polynucleotide. The nucleic
CC acids are useful as DNA vaccine against Streptococcus pneumoniae
CC infection (e.g. pneumonia). Nucleic acids derived from the S. pneumoniae
CC antigen nucleic acids are useful as probes for use in diagnostic methods
CC for detecting S. pneumoniae gene expression. The present sequence
CC represents an S. pneumoniae antigenic protein.
XX
SQ Sequence 270 AA;
Query Match 35.1%; Score 446.5; DB 7; Length 270;
Best Local Similarity 42.7%; Pred. No. 4.3e-41;
Matches 106; Conservative 41; Mismatches 94; Indels 7; Gaps 4;
QY 5 KYRADLVFLQGLAESREQAQRLIMAGKVTLTNNSTTIPRLLEKPGHKYPLESICSLIGV 64
DB 2 KERVDVLAYKQGLFETREQAKRGWAGLVVAVLNGE----RFDKPGEKIPDDTELKKG 57
QY 65 E-RFVSRGAYKLLTALDFFKIDVKSCICLDAGASTGGFTDCLLQHGASKYVYADVKGQL 123
DB 58 KLVYVSRGGKLEKALQVFDLSVDGATTIDIGASTGGFTDVM-LQNSAKLVFADVVGINQL 117
QY 124 HEKLYTNEQVINTEGVNLTASKDLIPEEVDILTIDVSFISLTILPSCIRWLKASGIII 183
DB 118 AWKLRQDPVRSMEQNFRYAEKTDPEQPSFASIDVSFISLSLILPALHRVLADQGVV 177
QY 184 ALIKPQFELYPDKI-KGVVVKETSLQYEAVEKIIHFQCSQELGLIFIGWVPSVIKPKGNQ 242
DB 178 ALVKPQFEAGREQIGKNGIIRDAKVHQNVLSTAMA-VEVGFSVLGLDFSPIQGGHNI 236
QY 243 EYLIYLKK 250
DB 237 EFLAYLKK 244
RESULT 5
ABB48304
ID ABB48304 standard; protein; 274 AA.
XX
AC ABB48304;
XX
DT 05-FEB-2002 (first entry)
XX
DE Listeria monocytogenes protein #1009.
XX
XX Antibacterial; gene therapy; vaccine; biosynthesis; biodegradation;
XX vitamin B12; bacterial infection; disease.
XX
XX Listeria monocytogenes.
XX
XX WO200177335-A2.
XX
XX 18-OCT-2001.
XX
XX 11-APR-2001; 2001WO-FR001118.
XX
XX 11-APR-2000; 2000FR-00004629.
XX
XX (INSP ) INST PASTEUR.
XX
XX Buchrieser C, Frangeul L, Couve E, Rusniok C, Fsihi H, Dehoux P;
XX Dussurget O, Chetouani F, Nedjari H, Glaser P, Kunst F, Cossart P;
XX Daniels J, Goebel W, Kreft J, Kuhn M, Ng E, Vazquez-Boland JA;
PI
```



```
Db      268 HGNEEYLLY 276
RESULT 7
ABP25556
ID      ABP25556 standard; protein; 275 AA.
XX
XX      ABP25556;
XX
XX      02-JUL-2002 (first entry)
XX
XX      Streptococcus polypeptide SEQ ID NO 288.
XX
XX      Streptococcus; GAS; GBS; group B streptococcus; Streptococcus agalactiae;
KW      group A streptococcus; Streptococcus pyogenes; antibacterial;
XX      antiinflammatory; infection; vaccine; meningitis; gene therapy.
XX
XX      Streptococcus pyogenes.
XX
XX      WO200234771-A2.
XX
XX      02-MAY-2002.
XX
XX      29-OCT-2001; 2001WO-GB004789.
XX
XX      27-OCT-2000; 2000GB-00026333.
PR      24-NOV-2000; 2000GB-00028727.
PR      07-MAR-2001; 2001GB-00005640.
XX
XX      (CHIR-) CHIRON SPA.
PA
PA      (GENO-) INST GENOMIC RES.
XX
XX      Telford J, Massignani V, Margarit Y RosI, Grandi G, Fraser C;
PI      Tettelin H;
XX
XX      WPI; 2002-352536/38.
DR      N-PSDB; ABN66187.
XX
XX      New Streptococcus protein for the treatment or prevention of infection or
PT      disease caused by Streptococcus bacteria, such as meningitis, and for
PT      detecting a compound that binds to the protein.
XX
XX      Claim 1; Page 3183; 4525pp; English.
XX
XX      The invention relates to a protein (ABP25413-ABP30895) from group B
CC      streptococcus/GBS (Streptococcus agalactiae) or group A streptococcus/GAS
CC      (Streptococcus pyogenes), comprising one of 5483 sequences (S1), given in
CC      the specification. The proteins have antibacterial and antiinflammatory
CC      activity. (I), nucleic acids encoding (I), ABN66044-ABN71526 and
CC      antibodies that bind (I) are used in the manufacture of medicaments for
CC      the treatment or prevention of infection or disease caused by
CC      Streptococcus bacteria, particularly S. agalactiae and S. pyogenes.
CC      Nucleic acids encoding (I) are used to detect Streptococcus in a
CC      biological sample. (I) is used to determine whether a compound binds to
CC      (I). A composition comprising (I) or a nucleic acid encoding (I), may be
CC      used as a vaccine or diagnostic composition. The disease caused by
CC      Streptococcus that is prevented or treated may be meningitis. Nucleic
CC      acid encoding (I) may be used to recombinantly produce (I) and may be
CC      used in gene therapy. Antibodies to (I) are used for affinity
CC      chromatography, immunoassays, and distinguishing/identifying
CC      Streptococcus proteins
XX
XX      Sequence 275 AA;
XX
Query Match      32.7%; Score 416; DB 5; Length 275;
Best Local Similarity 43.5%; Pred. No. 1.2e-37;
Matches 110; Conservative 37; Mismatches 90; Indels 16; Gaps 6;
QY      5 KVRDELVELVQGLAESREQAKRLIMAGKVTLTWNSTTIPLEKPGHKYPLESICSLIGV 64
Db      3 KERVVDVJAYKQGLFETREQARKGMAGLVSVINGQ----YDKPGDKIDGDTLKLKGE 58
QY      65 E-RFVSRGAYKLLTALDFFKIDVKSCICLDAGASTGGFTDCLLQHGASKVYVAIDVGRGQL 123
Db      59 KLVVSRGGLEKLEGLHVFQSVANQIGIDIGASTGGFTDMLQDQAKLVAVDVGNQL 118
QY      124 HEKLYTNEQVINIEGVNLRPTASKDLIPEVD-----ILTDVSPISLTLLPSCIRWLKA 178
Db      119 VVKLRQDPRVRSMEQYNFRYAQ----PEDFNEQGPVFASIDVSFISLSLILPALHNLSD 174
QY      179 SGIIIALIKPOFELYPPDKI-KKGWVKETSLOQYAVEKIIHFQCOSELGLIIGVVPSVIKG 237
Db      175 OGQVIALIKPOFAGREQIQKGIKXQIHEKVIQKVMDEA-SGYGFTVKGLDLDFSIQ 233
QY      238 PKGNQBYLYLKK 250
Db      234 GHGNIFFLAHLAK 246
XX
XX      RESULT 8
XX      ABP25555
XX      ID      ABP25555 standard; protein; 275 AA.
XX
XX      AC      ABP25555;
XX
XX      02-JUL-2002 (first entry)
XX
XX      Streptococcus polypeptide SEQ ID NO 288.
XX
XX      Streptococcus; GAS; GBS; group B streptococcus; Streptococcus agalactiae;
KW      group A streptococcus; Streptococcus pyogenes; antibacterial;
XX      antiinflammatory; infection; vaccine; meningitis; gene therapy.
XX
XX      Streptococcus agalactiae.
XX
XX      WO200234771-A2.
XX
XX      02-MAY-2002.
XX
XX      29-OCT-2001; 2001WO-GB004789.
XX
XX      27-OCT-2000; 2000GB-00026333.
PR      24-NOV-2000; 2000GB-00028727.
PR      07-MAR-2001; 2001GB-00005640.
XX
XX      (CHIR-) CHIRON SPA.
PA
PA      (GENO-) INST GENOMIC RES.
XX
XX      Telford J, Massignani V, Margarit Y RosI, Grandi G, Fraser C;
PI      Tettelin H;
XX
XX      WPI; 2002-352536/38.
DR      N-PSDB; ABN66186.
XX
XX      New Streptococcus protein for the treatment or prevention of infection or
PT      disease caused by Streptococcus bacteria, such as meningitis, and for
PT      detecting a compound that binds to the protein.
XX
XX      Claim 1; Page 3183; 4525pp; English.
XX
XX      The invention relates to a protein (ABP25413-ABP30895) from group B
CC      streptococcus/GBS (Streptococcus agalactiae) or group A streptococcus/GAS
CC      (Streptococcus pyogenes), comprising one of 5483 sequences (S1), given in
CC      the specification. The proteins have antibacterial and antiinflammatory
CC      activity. (I), nucleic acids encoding (I), ABN66044-ABN71526 and
CC      antibodies that bind (I) are used in the manufacture of medicaments for
CC      the treatment or prevention of infection or disease caused by
CC      Streptococcus bacteria, particularly S. agalactiae and S. pyogenes.
CC      Nucleic acids encoding (I) are used to detect Streptococcus in a
CC      biological sample. (I) is used to determine whether a compound binds to
CC      (I). A composition comprising (I) or a nucleic acid encoding (I), may be
CC      used as a vaccine or diagnostic composition. The disease caused by
CC      Streptococcus that is prevented or treated may be meningitis. Nucleic
CC      acid encoding (I) may be used to recombinantly produce (I) and may be
CC      used in gene therapy. Antibodies to (I) are used for affinity
CC      chromatography, immunoassays, and distinguishing/identifying
CC      Streptococcus proteins
XX
XX      Sequence 275 AA;
XX
```


CC infected mice, eliciting an immune response to T. hyodysenteriae
CC antigens. (Updated on 25-MAR-2003 to correct PF field.) (Updated on 25-
CC MAR-2003 to correct PR field.) (Updated on 16-OCT-2003 to standardise OS
CC field)
XX
SQ Sequence 240 AA;

Query Match 30.2%; Score 383.5; DB 2; Length 240;
Best Local Similarity 38.6%; Pred. No. 4.2e-34;
Matches 96; Conservative 41; Mismatches 93; Indels 19; Gaps 5

QY 6 VRADELFLQGLAESRQAKELINAGKYTLTNSTTIPRLLEKPKGHKYPLBSICSLIGVE 65
DB 1 MRLDEVHSEGYTSRSKAQDIILAGCVFVNG-----VKVTSKAHKIKDTDNIEVQNI 54
QY 66 RFVSRGAYKLLTALDFFKIDVKSCICLDAGASTGGFTDCLLQHGASKVYVIDVGKQLHE 125
DB 55 KYVSRAGEKLEKAFVEFGISVENKICLDIGASTGGFTDCLLKHGAKKYVALDVGHNLVY 114
QY 126 KLYTNEQVINIEGNLRTASKDLIPEEV-DILTIDVSFISITLILPSCIRMLKASGIIIA 184
DB 115 KLRNDRVVSIEDENAKDINKEMFNDEIPSVIVSDVSFISITKIAPIIIFKELNNLEFWWT 174
QY 185 LIRPQFELYDPDKIKKG-----VVKETSLQYEAEVKIHFQCOSELGLIFIGVVPVSIKCP 238
DB 175 LIRPQFAERGDVSKGGIIRDDILREKILN-NAISKII-----DCGFKEVNRNTISPIKGA 228
QY 239 KGNOEYLIY 247
DB 229 KGNIEYLAH 237

RESULT 11
AAR40751
ID AAR40751 standard; protein; 240 AA.
XX
AC AAR40751;
XX
DT 24-OCT-2003 (revised)
DT 27-AUG-2003 (revised)
DT 01-OCT-1993 (first entry)
XX
Sequence of haemolysin protein.
DE
XX Swine dysentery; vaccine; haemolysin protein.
KW
XX Brachyspira hyodysenteriae.
OS
XX
FN AU927224-A.
XX
PD 29-APR-1993.
XX
PF 22-OCT-1992; 92AU-00027224.
XX
PR 25-OCT-1991; 91EP-00202766.
PR 24-JUL-1992; 92EP-00202274.
XX
PA (DUIN) DUPHAR INT RES BV.
XX
PI Koopman MBH, Kusters JG, Muir SJ;
XX
XX WPI; 1993-188946/24.
DR N-PSDB; AAQ43504.
DR
XX
PT Polynucleotide encoding the haemolysin protein of Treponema
PT hyodysenteriae - used for obtaining polypeptide(s) for use in vaccines
PT for combating swine dysentery.
XX
PS Claim 5; Page 24-26; 41pp; English.
XX
CC DNA encoding TH polypeptides was isolated by cloning TH DNA and screening
CC for haemolytic colonies. AAQ43504 encodes haemolysin (AAR40751) and an
CC unknown protein (AAR38064). It is contained in cell line E. coli

[illegible]

PR 30-APR-1999;	99US-0132048P.	PR 21-JUL-1999;	99US-0145086P.
PR 30-APR-1999;	99US-0132407P.	PR 21-JUL-1999;	99US-0145088P.
PR 04-MAY-1999;	99US-0132484P.	PR 22-JUL-1999;	99US-0145085P.
PR 05-MAY-1999;	99US-0132485P.	PR 22-JUL-1999;	99US-0145087P.
PR 06-MAY-1999;	99US-0132486P.	PR 22-JUL-1999;	99US-0145089P.
PR 06-MAY-1999;	99US-0132487P.	PR 22-JUL-1999;	99US-0145192P.
PR 07-MAY-1999;	99US-0132863P.	PR 23-JUL-1999;	99US-0145145P.
PR 11-MAY-1999;	99US-0134256P.	PR 23-JUL-1999;	99US-0145218P.
PR 14-MAY-1999;	99US-0134218P.	PR 23-JUL-1999;	99US-0145224P.
PR 14-MAY-1999;	99US-0134219P.	PR 26-JUL-1999;	99US-0145276P.
PR 14-MAY-1999;	99US-0134221P.	PR 27-JUL-1999;	99US-0145913P.
PR 14-MAY-1999;	99US-0134370P.	PR 27-JUL-1999;	99US-0145918P.
PR 18-MAY-1999;	99US-0134768P.	PR 27-JUL-1999;	99US-0145919P.
PR 19-MAY-1999;	99US-0134941P.	PR 28-JUL-1999;	99US-0145951P.
PR 20-MAY-1999;	99US-0135124P.	PR 02-AUG-1999;	99US-0146386P.
PR 21-MAY-1999;	99US-0135353P.	PR 02-AUG-1999;	99US-0146388P.
PR 24-MAY-1999;	99US-0135629P.	PR 02-AUG-1999;	99US-0146389P.
PR 25-MAY-1999;	99US-0136021P.	PR 03-AUG-1999;	99US-0147038P.
PR 27-MAY-1999;	99US-0136392P.	PR 04-AUG-1999;	99US-0147204P.
PR 28-MAY-1999;	99US-0136782P.	PR 04-AUG-1999;	99US-0147302P.
PR 01-JUN-1999;	99US-0137222P.	PR 05-AUG-1999;	99US-0147192P.
PR 03-JUN-1999;	99US-0137528P.	PR 05-AUG-1999;	99US-0147260P.
PR 04-JUN-1999;	99US-0137502P.	PR 06-AUG-1999;	99US-0147303P.
PR 07-JUN-1999;	99US-0137724P.	PR 06-AUG-1999;	99US-0147416P.
PR 08-JUN-1999;	99US-0138094P.	PR 09-AUG-1999;	99US-0147493P.
PR 10-JUN-1999;	99US-0138540P.	PR 09-AUG-1999;	99US-0147935P.
PR 10-JUN-1999;	99US-0138847P.	PR 10-AUG-1999;	99US-0148171P.
PR 14-JUN-1999;	99US-0139119P.	PR 11-AUG-1999;	99US-0148319P.
PR 16-JUN-1999;	99US-0139452P.	PR 12-AUG-1999;	99US-0148341P.
PR 16-JUN-1999;	99US-0139453P.	PR 13-AUG-1999;	99US-0148565P.
PR 17-JUN-1999;	99US-0139492P.	PR 13-AUG-1999;	99US-0148684P.
PR 18-JUN-1999;	99US-0139454P.	PR 16-AUG-1999;	99US-0149368P.
PR 18-JUN-1999;	99US-0139455P.	PR 17-AUG-1999;	99US-0149175P.
PR 18-JUN-1999;	99US-0139456P.	PR 18-AUG-1999;	99US-0149426P.
PR 18-JUN-1999;	99US-0139457P.	PR 20-AUG-1999;	99US-0149722P.
PR 18-JUN-1999;	99US-0139458P.	PR 20-AUG-1999;	99US-0149723P.
PR 18-JUN-1999;	99US-0139459P.	PR 20-AUG-1999;	99US-0149929P.
PR 18-JUN-1999;	99US-0139460P.	PR 23-AUG-1999;	99US-0149902P.
PR 18-JUN-1999;	99US-0139461P.	PR 23-AUG-1999;	99US-0149930P.
PR 18-JUN-1999;	99US-0139462P.	PR 25-AUG-1999;	99US-0150566P.
PR 18-JUN-1999;	99US-0139463P.	PR 26-AUG-1999;	99US-0150884P.
PR 18-JUN-1999;	99US-0139750P.	PR 27-AUG-1999;	99US-0151065P.
PR 18-JUN-1999;	99US-0139763P.	PR 27-AUG-1999;	99US-0151066P.
PR 21-JUN-1999;	99US-0139817P.	PR 27-AUG-1999;	99US-0151080P.
PR 22-JUN-1999;	99US-0139899P.	PR 30-AUG-1999;	99US-0151303P.
PR 23-JUN-1999;	99US-0140353P.	PR 31-AUG-1999;	99US-0151438P.
PR 23-JUN-1999;	99US-0140354P.	PR 01-SEP-1999;	99US-0151930P.
PR 24-JUN-1999;	99US-0140695P.	PR 07-SEP-1999;	99US-0152363P.
PR 28-JUN-1999;	99US-0140823P.	PR 10-SEP-1999;	99US-0153070P.
PR 29-JUN-1999;	99US-0140991P.	PR 13-SEP-1999;	99US-0153758P.
PR 30-JUN-1999;	99US-0141287P.	PR 15-SEP-1999;	99US-0154018P.
PR 01-JUL-1999;	99US-0141842P.	PR 16-SEP-1999;	99US-0154039P.
PR 02-JUL-1999;	99US-0142154P.	PR 20-SEP-1999;	99US-0154779P.
PR 06-JUL-1999;	99US-0142055P.	PR 22-SEP-1999;	99US-0155139P.
PR 08-JUL-1999;	99US-0142390P.	PR 23-SEP-1999;	99US-0155486P.
PR 09-JUL-1999;	99US-0142803P.	PR 24-SEP-1999;	99US-0155659P.
PR 12-JUL-1999;	99US-0142920P.	PR 28-SEP-1999;	99US-0156458P.
PR 13-JUL-1999;	99US-0142977P.	PR 29-SEP-1999;	99US-0156596P.
PR 14-JUL-1999;	99US-0143542P.	PR 04-OCT-1999;	99US-0157117P.
PR 15-JUL-1999;	99US-0143624P.	PR 05-OCT-1999;	99US-0157753P.
PR 16-JUL-1999;	99US-0144005P.	PR 06-OCT-1999;	99US-0157865P.
PR 16-JUL-1999;	99US-0144085P.	PR 07-OCT-1999;	99US-0158029P.
PR 19-JUL-1999;	99US-0144086P.	PR 08-OCT-1999;	99US-0158232P.
PR 19-JUL-1999;	99US-0144325P.	PR 12-OCT-1999;	99US-0158369P.
PR 19-JUL-1999;	99US-0144331P.	PR 13-OCT-1999;	99US-0159293P.
PR 19-JUL-1999;	99US-0144332P.	PR 13-OCT-1999;	99US-0159294P.
PR 19-JUL-1999;	99US-0144333P.	PR 13-OCT-1999;	99US-0159295P.
PR 19-JUL-1999;	99US-0144334P.	PR 14-OCT-1999;	99US-0159329P.
PR 19-JUL-1999;	99US-0144335P.	PR 14-OCT-1999;	99US-0159330P.
PR 20-JUL-1999;	99US-0144352P.	PR 14-OCT-1999;	99US-0159331P.
PR 20-JUL-1999;	99US-0144632P.	PR 14-OCT-1999;	99US-0159637P.
PR 20-JUL-1999;	99US-0144884P.	PR 14-OCT-1999;	99US-0159638P.
PR 21-JUL-1999;	99US-0144814P.	PR 18-OCT-1999;	99US-0159584P.

PR	19-JUL-1999;	99US-0144331P.
PR	19-JUL-1999;	99US-0144332P.
PR	19-JUL-1999;	99US-0144333P.
PR	19-JUL-1999;	99US-0144334P.
PR	19-JUL-1999;	99US-0144335P.
PR	20-JUL-1999;	99US-01443352P.
PR	20-JUL-1999;	99US-01446332P.
PR	20-JUL-1999;	99US-0144884P.
PR	21-JUL-1999;	99US-0144814P.
PR	21-JUL-1999;	99US-0145086P.
PR	21-JUL-1999;	99US-0145088P.
PR	22-JUL-1999;	99US-0145085P.
PR	22-JUL-1999;	99US-0145087P.
PR	22-JUL-1999;	99US-0145089P.
PR	22-JUL-1999;	99US-0145192P.
PR	23-JUL-1999;	99US-0145145P.
PR	23-JUL-1999;	99US-0145218P.
PR	23-JUL-1999;	99US-0145224P.
PR	26-JUL-1999;	99US-0145276P.
PR	27-JUL-1999;	99US-0145913P.
PR	27-JUL-1999;	99US-0145918P.
PR	27-JUL-1999;	99US-0145919P.
PR	28-JUL-1999;	99US-0145951P.
PR	02-AUG-1999;	99US-0146386P.
PR	02-AUG-1999;	99US-0146388P.
PR	02-AUG-1999;	99US-0146389P.
PR	03-AUG-1999;	99US-0147038P.
PR	04-AUG-1999;	99US-0147204P.
PR	04-AUG-1999;	99US-0147302P.
PR	05-AUG-1999;	99US-0147192P.
PR	06-AUG-1999;	99US-0147260P.
PR	06-AUG-1999;	99US-0147303P.
PR	06-AUG-1999;	99US-0147416P.
PR	09-AUG-1999;	99US-0147493P.
PR	09-AUG-1999;	99US-0147935P.
PR	10-AUG-1999;	99US-0148171P.
PR	11-AUG-1999;	99US-0148319P.
PR	12-AUG-1999;	99US-0148341P.
PR	13-AUG-1999;	99US-0148565P.
PR	13-AUG-1999;	99US-0148684P.
PR	16-AUG-1999;	99US-0149368P.
PR	17-AUG-1999;	99US-0149375P.
PR	18-AUG-1999;	99US-0149426P.
PR	20-AUG-1999;	99US-0149722P.
PR	20-AUG-1999;	99US-0149723P.
PR	20-AUG-1999;	99US-0149929P.
PR	23-AUG-1999;	99US-0149902P.
PR	23-AUG-1999;	99US-0149930P.
PR	25-AUG-1999;	99US-0150566P.
PR	26-AUG-1999;	99US-0150884P.
PR	27-AUG-1999;	99US-0151065P.
PR	27-AUG-1999;	99US-0151066P.
PR	27-AUG-1999;	99US-0151080P.
PR	30-AUG-1999;	99US-0151303P.
PR	31-AUG-1999;	99US-0151438P.
PR	01-SEP-1999;	99US-0151930P.
PR	07-SEP-1999;	99US-0152363P.
PR	10-SEP-1999;	99US-0153070P.
PR	13-SEP-1999;	99US-0153758P.
PR	15-SEP-1999;	99US-0154018P.
PR	16-SEP-1999;	99US-0154039P.
PR	20-SEP-1999;	99US-0154779P.
PR	22-SEP-1999;	99US-0155139P.
PR	23-SEP-1999;	99US-0155486P.
PR	24-SEP-1999;	99US-0155659P.
PR	28-SEP-1999;	99US-0156458P.
PR	29-SEP-1999;	99US-0156596P.
PR	04-OCT-1999;	99US-0157117P.
PR	05-OCT-1999;	99US-0157753P.
PR	06-OCT-1999;	99US-0157865P.
PR	07-OCT-1999;	99US-0158029P.
PR	08-OCT-1999;	99US-0158032P.
PR	12-OCT-1999;	99US-0158369P.
PR	13-OCT-1999;	99US-0159293P.
PR	13-OCT-1999;	99US-0159294P.
PR	13-OCT-1999;	99US-0159295P.
PR	14-OCT-1999;	99US-0159329P.
PR	14-OCT-1999;	99US-0159330P.
PR	14-OCT-1999;	99US-0159331P.
PR	14-OCT-1999;	99US-0159637P.
PR	14-OCT-1999;	99US-0159638P.
PR	18-OCT-1999;	99US-0159584P.
PR	21-OCT-1999;	99US-0160741P.
PR	21-OCT-1999;	99US-0160767P.
PR	21-OCT-1999;	99US-0160768P.
PR	21-OCT-1999;	99US-0160770P.
PR	21-OCT-1999;	99US-0160814P.
PR	21-OCT-1999;	99US-0160815P.
PR	22-OCT-1999;	99US-0160980P.
PR	22-OCT-1999;	99US-0160981P.
PR	22-OCT-1999;	99US-0160989P.
PR	25-OCT-1999;	99US-0161404P.
PR	25-OCT-1999;	99US-0161405P.
PR	25-OCT-1999;	99US-0161406P.
PR	26-OCT-1999;	99US-0161359P.
PR	26-OCT-1999;	99US-0161360P.
PR	26-OCT-1999;	99US-0161361P.
PR	28-OCT-1999;	99US-0161920P.
PR	28-OCT-1999;	99US-0161992P.
PR	28-OCT-1999;	99US-0161993P.
PR	29-OCT-1999;	99US-0162142P.
Query Match 29.5%; Score 374.5; DB 3; Length 263;		
Best Local Similarity 37.1%; Pred. No. 4.9e-33;		
Matches 91; Conservative 49; Mismatches 90; Indels 15; Gaps 6;		
QY	5	KVRADLVFLQGLAESREQAKELIMAGKVTLTNNSTTIFLRLEKPGHKYPLESICSL--- 61
Db	16	KQRLDEACVERFNEYSRTLIQSWIMQGVLDGR-----RVIKAG--NPVSTTAIKIT 67
QY	62	IGVERFVSRGAYKLLTALDFKIDVKSCTCLDAGASTGFTDCLLOHGASKVVAIDVGKG 121
Db	68	AEVPKFCVCGGLKLEAALEKLDVDVSEKVLVDAGLSTGGFTDCLLRYGAAHVYGVVDVGX 127
QY	122	QLHEKLYTNEQVINIEGVNLRRTASKDLIPEEVDILTIDVYSFISLTLLPSCIRWLKASGI 181
Db	128	QVADKIRNDKKVTVIERTNLRVLPF--LPQKVDVTLDLISFISILKWPAINVMVKEDAT 185
QY	182	IIALIKPQFELYPDKI-KKGVVKETSLOEAVEKEIIHFQCSLGLIFIGVVPKVPKPGK 240
Db	186	LVTLYKPFQFEARRQVQVKGGIVRDPEVHOEVLKINGVE-RYGFTNKGFIESPICKADG 244
QY	241	NOEYL 245
Db	245	NIEFL 249
RESULT 14		
AAG23016		
ID	AAG23016	standard; protein; 301 AA.
XX	XX	
AC	AAG23016;	
XX	XX	
DT	17-OCT-2000	(first entry)
XX	XX	
DE	Arabidopsis thaliana	protein fragment SEQ ID NO: 26165.
XX	XX	
KW	Protein identification; signal transduction pathway; metabolic pathway;	
KW	hybridisation assay; genetic mapping; gene expression control; promoter;	
KW	termination sequence.	
XX	Arabidopsis thaliana.	
OS		
XX	EP1033405-A2.	
XX		
PD	06-SEP-2000.	

XX 25-FEB-2000; 2000EP-00301439.
PF 25-FEB-1999; 99US-0121825P.
XX 05-MAR-1999; 99US-0123180P.
PR 09-MAR-1999; 99US-0123548P.
PR 23-MAR-1999; 99US-0125788P.
PR 25-MAR-1999; 99US-0126264P.
PR 29-MAR-1999; 99US-0126785P.
PR 01-APR-1999; 99US-0127462P.
PR 06-APR-1999; 99US-0128234P.
PR 08-APR-1999; 99US-0128714P.
PR 16-APR-1999; 99US-0129845P.
PR 19-APR-1999; 99US-0130077P.
PR 21-APR-1999; 99US-0130449P.
PR 23-APR-1999; 99US-0130510P.
PR 28-APR-1999; 99US-0130891P.
PR 30-APR-1999; 99US-0131449P.
PR 30-APR-1999; 99US-0132048P.
PR 04-MAY-1999; 99US-0132407P.
PR 05-MAY-1999; 99US-0132484P.
PR 06-MAY-1999; 99US-0132485P.
PR 08-MAY-1999; 99US-0132486P.
PR 07-MAY-1999; 99US-0132487P.
PR 11-MAY-1999; 99US-0134256P.
PR 14-MAY-1999; 99US-0134218P.
PR 14-MAY-1999; 99US-0134219P.
PR 14-MAY-1999; 99US-0134221P.
PR 14-MAY-1999; 99US-0134370P.
PR 18-MAY-1999; 99US-0134768P.
PR 19-MAY-1999; 99US-0134941P.
PR 20-MAY-1999; 99US-0135124P.
PR 21-MAY-1999; 99US-0135353P.
PR 24-MAY-1999; 99US-0135629P.
PR 25-MAY-1999; 99US-0136021P.
PR 27-MAY-1999; 99US-0136393P.
PR 28-MAY-1999; 99US-0136782P.
PR 01-JUN-1999; 99US-0137222P.
PR 03-JUN-1999; 99US-0137528P.
PR 04-JUN-1999; 99US-0137502P.
PR 07-JUN-1999; 99US-0137724P.
PR 08-JUN-1999; 99US-0138094P.
PR 10-JUN-1999; 99US-0138540P.
PR 10-JUN-1999; 99US-0138847P.
PR 14-JUN-1999; 99US-0139119P.
PR 16-JUN-1999; 99US-0139452P.
PR 16-JUN-1999; 99US-0139453P.
PR 17-JUN-1999; 99US-0139492P.
PR 18-JUN-1999; 99US-0139454P.
PR 18-JUN-1999; 99US-0139455P.
PR 18-JUN-1999; 99US-0139456P.
PR 18-JUN-1999; 99US-0139457P.
PR 18-JUN-1999; 99US-0139458P.
PR 18-JUN-1999; 99US-0139459P.
PR 18-JUN-1999; 99US-0139460P.
PR 18-JUN-1999; 99US-0139461P.
PR 18-JUN-1999; 99US-0139462P.
PR 18-JUN-1999; 99US-0139463P.
PR 18-JUN-1999; 99US-0139750P.
PR 18-JUN-1999; 99US-0139763P.
PR 21-JUN-1999; 99US-0139817P.
PR 22-JUN-1999; 99US-0139899P.
PR 23-JUN-1999; 99US-0140353P.
PR 23-JUN-1999; 99US-0140354P.
PR 24-JUN-1999; 99US-0140695P.
PR 28-JUN-1999; 99US-0140823P.
PR 29-JUN-1999; 99US-0140991P.
PR 30-JUN-1999; 99US-0141287P.
PR 01-JUL-1999; 99US-0141842P.
PR 01-JUL-1999; 99US-0142154P.
PR 02-JUL-1999; 99US-0142055P.
PR 06-JUL-1999; 99US-0142390P.
PR 08-JUL-1999; 99US-0142803P.
PR 09-JUL-1999; 99US-0142920P.
PR 12-JUL-1999; 99US-0142977P.
PR 13-JUL-1999; 99US-0143542P.
PR 14-JUL-1999; 99US-0143624P.
PR 15-JUL-1999; 99US-0144005P.
PR 16-JUL-1999; 99US-0144085P.
PR 16-JUL-1999; 99US-0144086P.
PR 19-JUL-1999; 99US-0144325P.
PR 19-JUL-1999; 99US-0144331P.
PR 19-JUL-1999; 99US-0144332P.
PR 19-JUL-1999; 99US-0144333P.
PR 19-JUL-1999; 99US-0144334P.
PR 19-JUL-1999; 99US-0144335P.
PR 20-JUL-1999; 99US-0144352P.
PR 20-JUL-1999; 99US-0144632P.
PR 20-JUL-1999; 99US-0144884P.
PR 21-JUL-1999; 99US-0144814P.
PR 21-JUL-1999; 99US-0145086P.
PR 21-JUL-1999; 99US-0145088P.
PR 22-JUL-1999; 99US-0145085P.
PR 22-JUL-1999; 99US-0145087P.
PR 22-JUL-1999; 99US-0145089P.
PR 22-JUL-1999; 99US-0145192P.
PR 23-JUL-1999; 99US-0145145P.
PR 23-JUL-1999; 99US-0145218P.
PR 23-JUL-1999; 99US-0145224P.
PR 26-JUL-1999; 99US-0145276P.
PR 27-JUL-1999; 99US-0145913P.
PR 27-JUL-1999; 99US-0145918P.
PR 27-JUL-1999; 99US-0145919P.
PR 28-JUL-1999; 99US-0145951P.
PR 02-AUG-1999; 99US-0146386P.
PR 02-AUG-1999; 99US-0146388P.
PR 02-AUG-1999; 99US-0146389P.
PR 03-AUG-1999; 99US-0147038P.
PR 04-AUG-1999; 99US-0147204P.
PR 04-AUG-1999; 99US-0147302P.
PR 05-AUG-1999; 99US-0147192P.
PR 05-AUG-1999; 99US-0147260P.
PR 06-AUG-1999; 99US-0147303P.
PR 06-AUG-1999; 99US-0147416P.
PR 09-AUG-1999; 99US-0147493P.
PR 09-AUG-1999; 99US-0147935P.
PR 10-AUG-1999; 99US-0148171P.
PR 11-AUG-1999; 99US-0148319P.
PR 12-AUG-1999; 99US-0148341P.
PR 13-AUG-1999; 99US-0148565P.
PR 13-AUG-1999; 99US-0148684P.
PR 16-AUG-1999; 99US-0149368P.
PR 17-AUG-1999; 99US-0149175P.
PR 18-AUG-1999; 99US-0149426P.
PR 20-AUG-1999; 99US-0149722P.
PR 20-AUG-1999; 99US-0149723P.
PR 20-AUG-1999; 99US-0149929P.
PR 23-AUG-1999; 99US-0149902P.
PR 23-AUG-1999; 99US-0149930P.
PR 25-AUG-1999; 99US-0150566P.
PR 26-AUG-1999; 99US-0150884P.
PR 27-AUG-1999; 99US-0151065P.
PR 27-AUG-1999; 99US-0151066P.
PR 27-AUG-1999; 99US-0151080P.
PR 30-AUG-1999; 99US-0151303P.
PR 31-AUG-1999; 99US-0151438P.
PR 01-SEP-1999; 99US-0151930P.
PR 07-SEP-1999; 99US-0152363P.
PR 10-SEP-1999; 99US-0153070P.
PR 13-SEP-1999; 99US-0153758P.
PR 15-SEP-1999; 99US-0154018P.
PR 16-SEP-1999; 99US-0154039P.
PR 20-SEP-1999; 99US-0154779P.
PR 22-SEP-1999; 99US-0155139P.
PR 23-SEP-1999; 99US-0155486P.

PR	24-SEP-1999;	99US-0155659P.
PR	28-SEP-1999;	99US-0156458P.
PR	29-SEP-1999;	99US-0156596P.
PR	04-OCT-1999;	99US-0157117P.
PR	05-OCT-1999;	99US-0157753P.
PR	06-OCT-1999;	99US-0157865P.
PR	07-OCT-1999;	99US-0158029P.
PR	08-OCT-1999;	99US-0158232P.
PR	12-OCT-1999;	99US-0158369P.
PR	13-OCT-1999;	99US-0159293P.
PR	13-OCT-1999;	99US-0159294P.
PR	13-OCT-1999;	99US-0159295P.
PR	14-OCT-1999;	99US-0159329P.
PR	14-OCT-1999;	99US-0159330P.
PR	14-OCT-1999;	99US-0159331P.
PR	14-OCT-1999;	99US-0159637P.
PR	14-OCT-1999;	99US-0159638P.
PR	18-OCT-1999;	99US-0159584P.
PR	21-OCT-1999;	99US-0160741P.
PR	21-OCT-1999;	99US-0160767P.
PR	21-OCT-1999;	99US-0160768P.
PR	21-OCT-1999;	99US-0160770P.
PR	21-OCT-1999;	99US-0160814P.
PR	21-OCT-1999;	99US-0160815P.
PR	22-OCT-1999;	99US-0160980P.
PR	22-OCT-1999;	99US-0160981P.
PR	22-OCT-1999;	99US-0160989P.
PR	25-OCT-1999;	99US-0161404P.
PR	25-OCT-1999;	99US-0161405P.
PR	25-OCT-1999;	99US-0161406P.
PR	26-OCT-1999;	99US-0161359P.
PR	26-OCT-1999;	99US-0161360P.
PR	26-OCT-1999;	99US-0161361P.
PR	28-OCT-1999;	99US-0161920P.
PR	28-OCT-1999;	99US-0161992P.
PR	28-OCT-1999;	99US-0161993P.
PR	29-OCT-1999;	99US-0162142P.

Query Match

29.5%; Score 374.5; DB 3; Length 301;

Best Local Similarity 37.1%; Pred. No. 5.9e-33;

Matches 91; Conservative 49; Mismatches 90; Indels 15; Gaps 6;

QY	5	KVRDELVPFQGLAESREQAKRLIMAGKVTLLNNSTTIFLRLEKPGHKYPLSICSL---	61
Db	54	KQRLDEACVERENEYSRTLIQSWIMOGKVLVDGR-----RVIKAG--MPVSTTAAIKIT	105
QY	62	IGVERFVSQAYKLLTALDFPKLDVKSCLDAGASTGGTDCLLQHGASKVYADIVGKG	121
Db	106	AEVPFVCRGGGLKLEALEKLDVDSEKVVLDAGLSTGGFTDCLLRGAHVGVGVG	165
QY	122	QLHEKLYTNEQVINIEGVNLTASKDLIPEVDILTIDVSFISLTLILPSCIRWLKASGI	181
Db	166	QVADKIRNDKQVTVIBRTNLYLPG--LPQKVDVTVLDSLFSILKVMFALNNVMKEDAT	223
QY	182	ITALIKPQELYPDKI-KXGVVKETSLOYEAVEKEIHFQCOSELGLIFIGVVPSEVIKPGKG	240
Db	224	LVTLVKPFQEARQQVQGGKGIIVDPDVHGVLEKIINGVE-RYGFTNKGFIESPICKGADG	282
QY	241	NOEYL 245	
Db	283	NIEFL 287	

RESULT 15

AAG51201

ID AAG51201 standard; protein; 303 AA.

XX AC AAG51201;

XX AC AAG51201;

18-OCT-2000 (first entry)

Arabidopsis thaliana protein fragment SEQ ID NO: 64959.

XX

DE

KW	Protein identification; signal transduction pathway; metabolic pathway; hybridisation assay; genetic mapping; gene expression control; promoter; termination sequence.
KW	Arabidopsis thaliana.
XX	EP1033405-A2.
XX	06-SEP-2000.
XX	25-FEB-2000; 2000EP-00301439.
XX	25-FEB-1999; 99US-0121825P.
PR	05-MAR-1999; 99US-0123180P.
PR	09-MAR-1999; 99US-0123548P.
PR	23-MAR-1999; 99US-0125788P.
PR	25-MAR-1999; 99US-0126264P.
PR	29-MAR-1999; 99US-0126785P.
PR	01-APR-1999; 99US-0127462P.
PR	06-APR-1999; 99US-0128234P.
PR	08-APR-1999; 99US-0128714P.
PR	16-APR-1999; 99US-0129845P.
PR	19-APR-1999; 99US-0130077P.
PR	21-APR-1999; 99US-0130449P.
PR	23-APR-1999; 99US-0130510P.
PR	23-APR-1999; 99US-0130891P.
PR	28-APR-1999; 99US-0131449P.
PR	30-APR-1999; 99US-0132048P.
PR	30-APR-1999; 99US-0132407P.
PR	04-MAY-1999; 99US-0132484P.
PR	05-MAY-1999; 99US-0132485P.
PR	06-MAY-1999; 99US-0132486P.
PR	06-MAY-1999; 99US-0132487P.
PR	07-MAY-1999; 99US-0132863P.
PR	11-MAY-1999; 99US-0134256P.
PR	14-MAY-1999; 99US-0134218P.
PR	14-MAY-1999; 99US-0134219P.
PR	14-MAY-1999; 99US-0134221P.
PR	14-MAY-1999; 99US-0134370P.
PR	18-MAY-1999; 99US-0134768P.
PR	19-MAY-1999; 99US-0134941P.
PR	20-MAY-1999; 99US-0135124P.
PR	21-MAY-1999; 99US-0135353P.
PR	24-MAY-1999; 99US-0135629P.
PR	25-MAY-1999; 99US-0136021P.
PR	27-MAY-1999; 99US-0136392P.
PR	28-MAY-1999; 99US-0136782P.
PR	01-JUN-1999; 99US-0137222P.
PR	03-JUN-1999; 99US-0137528P.
PR	04-JUN-1999; 99US-0137502P.
PR	07-JUN-1999; 99US-0137724P.
PR	08-JUN-1999; 99US-0138094P.
PR	10-JUN-1999; 99US-0138540P.
PR	10-JUN-1999; 99US-0138847P.
PR	14-JUN-1999; 99US-0139119P.
PR	16-JUN-1999; 99US-0139452P.
PR	16-JUN-1999; 99US-0139453P.
PR	17-JUN-1999; 99US-0139492P.
PR	18-JUN-1999; 99US-0139454P.
PR	18-JUN-1999; 99US-0139455P.
PR	18-JUN-1999; 99US-0139456P.
PR	18-JUN-1999; 99US-0139457P.
PR	18-JUN-1999; 99US-0139458P.
PR	18-JUN-1999; 99US-0139459P.
PR	18-JUN-1999; 99US-0139460P.
PR	18-JUN-1999; 99US-0139461P.
PR	18-JUN-1999; 99US-0139462P.
PR	18-JUN-1999; 99US-0139463P.
PR	18-JUN-1999; 99US-0139750P.
PR	18-JUN-1999; 99US-0139763P.
PR	21-JUN-1999; 99US-0139817P.
PR	22-JUN-1999; 99US-0139899P.
PR	23-JUN-1999; 99US-0140353P.

PR	23-JUN-1999;	99US-0140354P.	PR	01-SEP-1999;	99US-0151930P.
PR	24-JUN-1999;	99US-0140695P.	PR	07-SEP-1999;	99US-0152363P.
PR	28-JUN-1999;	99US-0140823P.	PR	10-SEP-1999;	99US-0153070P.
PR	29-JUN-1999;	99US-0140991P.	PR	13-SEP-1999;	99US-0153758P.
PR	30-JUN-1999;	99US-0141287P.	PR	15-SEP-1999;	99US-0154018P.
PR	01-JUL-1999;	99US-0141842P.	PR	16-SEP-1999;	99US-0154039P.
PR	02-JUL-1999;	99US-0142154P.	PR	20-SEP-1999;	99US-0154779P.
PR	03-JUL-1999;	99US-0142055P.	PR	22-SEP-1999;	99US-0155139P.
PR	06-JUL-1999;	99US-0142390P.	PR	23-SEP-1999;	99US-0155486P.
PR	08-JUL-1999;	99US-0142803P.	PR	24-SEP-1999;	99US-0155659P.
PR	09-JUL-1999;	99US-0142920P.	PR	28-SEP-1999;	99US-0156458P.
PR	12-JUL-1999;	99US-0142977P.	PR	29-SEP-1999;	99US-0156596P.
PR	13-JUL-1999;	99US-0143542P.	PR	04-OCT-1999;	99US-0157117P.
PR	14-JUL-1999;	99US-0143624P.	PR	05-OCT-1999;	99US-0157753P.
PR	15-JUL-1999;	99US-0144005P.	PR	06-OCT-1999;	99US-0157865P.
PR	16-JUL-1999;	99US-0144085P.	PR	07-OCT-1999;	99US-0158029P.
PR	18-JUL-1999;	99US-0144086P.	PR	08-OCT-1999;	99US-0158232P.
PR	19-JUL-1999;	99US-0144331P.	PR	12-OCT-1999;	99US-0158369P.
PR	19-JUL-1999;	99US-0144332P.	PR	13-OCT-1999;	99US-0159293P.
PR	19-JUL-1999;	99US-0144333P.	PR	13-OCT-1999;	99US-0159294P.
PR	19-JUL-1999;	99US-0144334P.	PR	14-OCT-1999;	99US-0159295P.
PR	20-JUL-1999;	99US-0144335P.	PR	14-OCT-1999;	99US-0159329P.
PR	20-JUL-1999;	99US-0144352P.	PR	14-OCT-1999;	99US-0159330P.
PR	20-JUL-1999;	99US-0144632P.	PR	14-OCT-1999;	99US-0159331P.
PR	21-JUL-1999;	99US-0144884P.	PR	14-OCT-1999;	99US-0159637P.
PR	21-JUL-1999;	99US-0144814P.	PR	18-OCT-1999;	99US-0159638P.
PR	21-JUL-1999;	99US-0145086P.	PR	18-OCT-1999;	99US-0159584P.
PR	21-JUL-1999;	99US-0145088P.	PR	21-OCT-1999;	99US-0160741P.
PR	22-JUL-1999;	99US-0145085P.	PR	21-OCT-1999;	99US-0160767P.
PR	22-JUL-1999;	99US-0145087P.	PR	21-OCT-1999;	99US-0160768P.
PR	22-JUL-1999;	99US-0145089P.	PR	21-OCT-1999;	99US-0160770P.
PR	22-JUL-1999;	99US-0145192P.	PR	21-OCT-1999;	99US-0160814P.
PR	23-JUL-1999;	99US-0145145P.	PR	21-OCT-1999;	99US-0160815P.
PR	23-JUL-1999;	99US-0145218P.	PR	22-OCT-1999;	99US-0160880P.
PR	23-JUL-1999;	99US-0145224P.	PR	22-OCT-1999;	99US-0160981P.
PR	26-JUL-1999;	99US-0145276P.	PR	22-OCT-1999;	99US-0160989P.
PR	27-JUL-1999;	99US-0145913P.	PR	22-OCT-1999;	99US-0161404P.
PR	27-JUL-1999;	99US-0145918P.	PR	25-OCT-1999;	99US-0161405P.
PR	27-JUL-1999;	99US-0145919P.	PR	25-OCT-1999;	99US-0161406P.
PR	28-JUL-1999;	99US-0145951P.	PR	26-OCT-1999;	99US-0161359P.
PR	02-AUG-1999;	99US-0146386P.	PR	26-OCT-1999;	99US-0161360P.
PR	02-AUG-1999;	99US-0146388P.	PR	26-OCT-1999;	99US-0161361P.
PR	02-AUG-1999;	99US-0146389P.	PR	28-OCT-1999;	99US-0161920P.
PR	03-AUG-1999;	99US-0147038P.	PR	28-OCT-1999;	99US-0161922P.
PR	04-AUG-1999;	99US-0147204P.	PR	28-OCT-1999;	99US-0161993P.
PR	04-AUG-1999;	99US-0147302P.	PR	29-OCT-1999;	99US-0162142P.
PR	05-AUG-1999;	99US-0147192P.	Query Match 29.5%; Score 374.5; DB 3; Length 303;		
PR	05-AUG-1999;	99US-0147260P.	Best Local Similarity 37.1%; Pred. No. 6e-33;		
PR	06-AUG-1999;	99US-0147303P.	Matches 91; Conservative 49; Mismatches 90; Indels 15; Gaps 6;		
PR	06-AUG-1999;	99US-0147416P.			
PR	09-AUG-1999;	99US-0147493P.	QY	5	KVRADELVFLQLAESRQAKLIMAGKVTLTNNSTTILRLKPGHKYPLESICSL---61
PR	09-AUG-1999;	99US-0147935P.	Db	56	KQRLDEACVREFNEYSRTLIQSWIMQGVLDGR-----RVIKAG--MPVSTTAAIKIT107
PR	10-AUG-1999;	99US-0148171P.	QY	62	IGVERFVSRGAYKLLTALDFFKIDVKSCICLDAGASTGGTDCLLQHGASKVVAIDVGKG121
PR	11-AUG-1999;	99US-0148319P.	Db	108	AEVPKFCVCGGLKLEALEKLDVDSEKVLDAGSTGGTDCLLRYGAHVGVVDVGYG167
PR	12-AUG-1999;	99US-0148341P.	QY	122	QLHEKLYTNEQVINIEGVNLRTASKDLIPEEVDILTIDVSFISLTILPSCIRWLKASGI181
PR	13-AUG-1999;	99US-0148565P.	Db	168	QVADKIRNDKKVTVIERTNLRYPG--LPQKVDVVTLDLSFISILKVMFPAIXNMKEDAT225
PR	16-AUG-1999;	99US-0149368P.	QY	182	IYALIKPQPELYPDKI-KGVVKETSLQYEAVEKIHFCQSELGLIFIGVPSVIKPGKG240
PR	17-AUG-1999;	99US-0149175P.	Db	226	LVTLVKQPQFARRQQVQGGIVRDPVHQVELEKIINGVE-RYGFTNKGFIESPICKGADG284
PR	18-AUG-1999;	99US-0149426P.	QY	241	NOEYL245
PR	20-AUG-1999;	99US-0149722P.	Db	285	NIEFL289
PR	20-AUG-1999;	99US-0149723P.			
PR	20-AUG-1999;	99US-0149929P.			
PR	23-AUG-1999;	99US-0149902P.			
PR	23-AUG-1999;	99US-0149930P.			
PR	25-AUG-1999;	99US-0150566P.			
PR	26-AUG-1999;	99US-0150884P.			
PR	27-AUG-1999;	99US-0151065P.			
PR	27-AUG-1999;	99US-0151066P.			
PR	27-AUG-1999;	99US-0151080P.			
PR	30-AUG-1999;	99US-0151303P.			
PR	31-AUG-1999;	99US-0151438P.			

RESULT 16

PR 16-JUN-1999; 99US-0139452P.
PR 16-JUN-1999; 99US-0139453P.
PR 17-JUN-1999; 99US-0139492P.
PR 18-JUN-1999; 99US-0139454P.
PR 18-JUN-1999; 99US-0139455P.
PR 18-JUN-1999; 99US-0139456P.
PR 18-JUN-1999; 99US-0139457P.
PR 18-JUN-1999; 99US-0139458P.
PR 18-JUN-1999; 99US-0139459P.
PR 18-JUN-1999; 99US-0139460P.
PR 18-JUN-1999; 99US-0139461P.
PR 18-JUN-1999; 99US-0139462P.
PR 18-JUN-1999; 99US-0139463P.
PR 18-JUN-1999; 99US-0139750P.
PR 18-JUN-1999; 99US-0139763P.
PR 21-JUN-1999; 99US-0139817P.
PR 22-JUN-1999; 99US-0139899P.
PR 23-JUN-1999; 99US-0140353P.
PR 23-JUN-1999; 99US-0140354P.
PR 24-JUN-1999; 99US-0140695P.
PR 28-JUN-1999; 99US-0140823P.
PR 29-JUN-1999; 99US-0140991P.
PR 30-JUN-1999; 99US-0141287P.
PR 01-JUL-1999; 99US-0141642P.
PR 01-JUL-1999; 99US-0142154P.
PR 02-JUL-1999; 99US-0142055P.
PR 06-JUL-1999; 99US-0142390P.
PR 08-JUL-1999; 99US-0142803P.
PR 09-JUL-1999; 99US-0142920P.
PR 12-JUL-1999; 99US-0142977P.
PR 13-JUL-1999; 99US-0143542P.
PR 14-JUL-1999; 99US-0143624P.
PR 15-JUL-1999; 99US-0144005P.
PR 16-JUL-1999; 99US-0144085P.
PR 16-JUL-1999; 99US-0144086P.
PR 19-JUL-1999; 99US-0144325P.
PR 19-JUL-1999; 99US-0144331P.
PR 19-JUL-1999; 99US-0144332P.
PR 19-JUL-1999; 99US-0144333P.
PR 19-JUL-1999; 99US-0144334P.
PR 19-JUL-1999; 99US-0144335P.
PR 20-JUL-1999; 99US-0144352P.
PR 20-JUL-1999; 99US-0144632P.
PR 20-JUL-1999; 99US-0144884P.
PR 21-JUL-1999; 99US-0144814P.
PR 21-JUL-1999; 99US-0145086P.
PR 21-JUL-1999; 99US-0145088P.
PR 22-JUL-1999; 99US-0145085P.
PR 22-JUL-1999; 99US-0145087P.
PR 22-JUL-1999; 99US-0145089P.
PR 22-JUL-1999; 99US-0145192P.
PR 23-JUL-1999; 99US-0145145P.
PR 23-JUL-1999; 99US-0145218P.
PR 23-JUL-1999; 99US-0145224P.
PR 26-JUL-1999; 99US-0145276P.
PR 27-JUL-1999; 99US-0145913P.
PR 27-JUL-1999; 99US-0145918P.
PR 27-JUL-1999; 99US-0145919P.
PR 28-JUL-1999; 99US-0145951P.
PR 02-AUG-1999; 99US-0146386P.
PR 02-AUG-1999; 99US-0146388P.
PR 02-AUG-1999; 99US-0146389P.
PR 03-AUG-1999; 99US-0147038P.
PR 04-AUG-1999; 99US-0147204P.
PR 04-AUG-1999; 99US-0147302P.
PR 05-AUG-1999; 99US-0147192P.
PR 05-AUG-1999; 99US-0147260P.
PR 06-AUG-1999; 99US-0147303P.
PR 06-AUG-1999; 99US-0147416P.
PR 09-AUG-1999; 99US-0147493P.
PR 09-AUG-1999; 99US-0147935P.
PR 10-AUG-1999; 99US-0148171P.
PR 11-AUG-1999; 99US-0148319P.

PR 12-AUG-1999; 99US-0148341P.
PR 13-AUG-1999; 99US-0148565P.
PR 13-AUG-1999; 99US-0148684P.
PR 16-AUG-1999; 99US-0149368P.
PR 17-AUG-1999; 99US-0149175P.
PR 18-AUG-1999; 99US-0149428P.
PR 20-AUG-1999; 99US-0149722P.
PR 20-AUG-1999; 99US-0149723P.
PR 20-AUG-1999; 99US-0149929P.
PR 23-AUG-1999; 99US-0149902P.
PR 23-AUG-1999; 99US-0149930P.
PR 25-AUG-1999; 99US-0150566P.
PR 26-AUG-1999; 99US-0150884P.
PR 27-AUG-1999; 99US-0151065P.
PR 27-AUG-1999; 99US-0151066P.
PR 27-AUG-1999; 99US-0151080P.
PR 30-AUG-1999; 99US-0151303P.
PR 31-AUG-1999; 99US-0151438P.
PR 01-SEP-1999; 99US-0151930P.
PR 07-SEP-1999; 99US-0152363P.
PR 10-SEP-1999; 99US-0153070P.
PR 13-SEP-1999; 99US-0153758P.
PR 15-SEP-1999; 99US-0154018P.
PR 16-SEP-1999; 99US-0154039P.
PR 20-SEP-1999; 99US-0154779P.
PR 22-SEP-1999; 99US-0155139P.
PR 23-SEP-1999; 99US-0155486P.
PR 24-SEP-1999; 99US-0155659P.
PR 28-SEP-1999; 99US-0156458P.
PR 29-SEP-1999; 99US-0156596P.
PR 04-OCT-1999; 99US-0157117P.
PR 05-OCT-1999; 99US-0157753P.
PR 06-OCT-1999; 99US-0157865P.
PR 07-OCT-1999; 99US-0158029P.
PR 08-OCT-1999; 99US-0158232P.
PR 12-OCT-1999; 99US-0158369P.
PR 13-OCT-1999; 99US-0159293P.
PR 13-OCT-1999; 99US-0159294P.
PR 13-OCT-1999; 99US-0159295P.
PR 14-OCT-1999; 99US-0159329P.
PR 14-OCT-1999; 99US-0159330P.
PR 14-OCT-1999; 99US-0159331P.
PR 14-OCT-1999; 99US-0159637P.
PR 14-OCT-1999; 99US-0159638P.
PR 18-OCT-1999; 99US-0159584P.
PR 21-OCT-1999; 99US-0160741P.
PR 21-OCT-1999; 99US-0160767P.
PR 21-OCT-1999; 99US-0160770P.
PR 21-OCT-1999; 99US-0160814P.
PR 21-OCT-1999; 99US-0160815P.
PR 22-OCT-1999; 99US-0160980P.
PR 22-OCT-1999; 99US-0160981P.
PR 22-OCT-1999; 99US-0160989P.
PR 25-OCT-1999; 99US-0161404P.
PR 25-OCT-1999; 99US-0161405P.
PR 25-OCT-1999; 99US-0161406P.
PR 26-OCT-1999; 99US-0161359P.
PR 26-OCT-1999; 99US-0161360P.
PR 26-OCT-1999; 99US-0161361P.
PR 28-OCT-1999; 99US-0161920P.
PR 28-OCT-1999; 99US-0161992P.
PR 28-OCT-1999; 99US-0161993P.
PR 29-OCT-1999; 99US-0162142P.

Query Match 29.5%; Score 374.5; DB 3; Length 335;

Best Local Similarity 37.1%; Pred. No. 7e-33;

Matches 91; Conservative 49; Mismatches 90; Indels 15; Gaps 6;

QY 5 KVRADLVFLQGLAESREQAKLIMAGKVTLTNNSTTIFLRLEKPGHKYPLESICSL---

Db 88 KQRLDEACVERFNEYSRTLIQSWIMQGVLDGR-----RVIKAG--MPVSTAAIKIT 139

Query Match 28.6%; Score 363; DB 5; Length 243;
Best Local Similarity 40.4%; Pred. No. 8.4e-32;
Matches 91; Conservative 40; Mismatches 86; Indels 8; Gaps 5;
QY 29 MAGKVTLTNNSTTIPURLEKPGHKYPLESICSLIGVE-RFVSRGAYKLTALDFFKIDVK 87
DB 1 MAGLVVDSKSGE---RYDKPGQIDDDGTELRKLGKRYVSRGGLKLEKALKEFHEIN 56
QY 88 SCICLDAGASTGFTDCLLOHGASKYVADVGQOLHEKLYTNEQVNIETGVNLRFA-SK 146
DB 57 GKTCLDIGSGTGFTDVMLONGAKLYALDVGNTQAWKIRSDERVVVMEQNFNRNAVLA 116
QY 147 DLIPBEVDLTIDVSFISLTILPSCIRWLKASGIIIALIKQFELYPDKI-KKGVVKET 205
DB 117 DEQGEPSFTSDVSFISLIDLILPPLYEILEKNGEVAALIKQFEAGRGVQVGNRIIDP 176
QY 206 SIQYEAWEKIIHFCQSELGLIFIGVVPSVIKPGKNOEVLIIYKK 250
DB 177 KVHQMTIEKVLK-TATQLGFSVKGLTFSPIKGAGNVFELVLLK 220

RESULT 20
ADB09362
ID ADB09362 standard; protein; 268 AA.
AC ADB09362;
DT 20-NOV-2003 (first entry)
DE Alloiooccus otitis antigenic protein SEQ ID NO:3302.
XX
XX Alloiooccus otitidis; antigenic protein; immunogenic; immunisation;
KW Alloiooccus otitidis; antigenic protein; immunogenic; immunisation;
KW gene therapy; Gram-positive bacterium; infection.
XX
OS Alloiooccus otitis.
XX
XX WO2003048304-A2.
XX
XX 12-JUN-2003.
XX
XX 25-NOV-2002; 2002WO-US036123.
XX
XX 29-NOV-2001; 2001US-0333777P.
XX
XX 18-NOV-2002; 2002US-0426742P.
XX
XX (AMHP) WYETH HOLDINGS CORP.
XX
XX Fletcher LD, Mcmichael JC, Russell DP, Zagursky RJ;
XX WPI; 2003-505284/47.
XX N-PSDB; ADB09361.
XX
XX New Alloiooccus otitidis polynucleotides and polypeptides, useful for
PT treating and diagnosing diseases, drug screening assays and monitoring of
PT effects during drug clinical trials.
XX
XX Claim 33; SEQ ID NO 3302; 1019pp; English.

The present invention describes an isolated polynucleotide (1) of
CC Alloiooccus otitidis genomic DNA, which encodes an antigenic protein.
CC Alloiooccus otitidis is a Gram-positive bacterium. Also described: (1)
CC an isolated polypeptide that is encoded by the polynucleotide (1); (2) an
CC expression vector comprising the novel isolated polynucleotide (1), its
CC complement, degenerate variant or fragment; (3) a genetically engineered
CC host cell, transfected, transformed or infected with the vector of (2);
CC (4) an antibody specific for the polypeptide of (1); (5) an immunogenic
CC composition comprising the polypeptide, its complement, biological
CC equivalent or fragment, or the polynucleotide that is comprised in the
CC expression vector; (6) a pharmaceutical composition comprising the
CC polypeptide of (1) and a carrier; (7) a protein chip comprising an array
CC of the polypeptides of (1), their biological equivalent or fragment; (8)
CC immunising against Alloiooccus otitidis by administering to a host the
CC immunogenic composition; (9) detecting and/or identifying Alloiooccus

CC otitidis in the biological sample; (10) a kit comprising a container
CC containing the novel polynucleotide, its degenerate variant or fragment,
CC or the antibody of (4); and (11) producing a polypeptide by culturing the
CC genetically engineered host cell under conditions suitable to produce the
CC polypeptide from the culture. (1) can be used in gene therapy. The
CC polynucleotides, polypeptides, antibodies and compositions of the present
CC invention can be used for treating and diagnosing diseases, drug
CC screening assays and monitoring of effects during drug clinical trials.
CC The polynucleotides are useful for expressing and detecting Alloiooccus
CC otitidis. The present sequence represents an Alloiooccus otitidis
CC antigen protein from the present invention.

XX
SQ Sequence 268 AA;
Query Match 28.2%; Score 358; DB 6; Length 268;
Best Local Similarity 35.6%; Pred. No. 3.5e-31;
Matches 95; Conservative 45; Mismatches 87; Indels 40; Gaps 8;
QY 1 MAKHKVRADELVFLQGLAESRQAKRLIMAGKV-TJTNNSTTIP-----LRLEKPG 50
DB 1 MAK-KIRADQAMLDQDLDDLXEAQALIMAGRYVTYVKESQVPTPGYQVKEGTQLYIKSPD 59
QY 51 HKYPLESICSLIGVERFVSRGAYKLTALDFFKIDVKSCICLDAGASTGFTDCLLOHGA 110
DB 60 HPY-----VSRGGLKQKQALDSFDLPFAGKILIDIGSGTGFTDVALRHGA 105
QY 111 SKVYADVGKQOLHEKLYTNEQVNIETGVNLRFA-SKDLIPBEVDLTIDVSFISLTIL 169
DB 106 GLVYALDVGNTQVWVSLRQDDRVIVMEQTNFKDCKLSDFRAGQPDLAVMDSFISLYPIL 165
QY 170 PSCIRWLKASGIIIALIKQFELYPDKI-K-KGVKETSLOYEAVEKI-----IHFQSE 222
DB 166 ENLAQILPPGHLAVALIKQFEADKEDVQPGSLVLDPPGVHRKVLLEATTAFMTKLHYC--- 222
QY 223 LGLIFIGVVPSVIKPGKNOEVLIIYK 249
DB 223 ----LLGLLASPIKGGKNIETALLR 245

RESULT 21
ADB09364
ID ADB09364 standard; protein; 271 AA.
XX
XX ADB09364;
XX
XX 20-NOV-2003 (first entry)
XX
XX Alloiooccus otitis antigenic protein SEQ ID NO:3304.
XX
XX Alloiooccus otitidis; antigenic protein; immunogenic; immunisation;
KW gene therapy; Gram-positive bacterium; infection.
XX
XX Alloiooccus otitis.
XX
XX WO2003048304-A2.
XX
XX 12-JUN-2003.
XX
XX 25-NOV-2002; 2002WO-US036123.
XX
XX 29-NOV-2001; 2001US-0333777P.
XX
XX 18-NOV-2002; 2002US-0426742P.
XX
XX (AMHP) WYETH HOLDINGS CORP.
XX
XX Fletcher LD, Mcmichael JC, Russell DP, Zagursky RJ;
XX WPI; 2003-505284/47.
XX N-PSDB; ADB09363.
XX
XX New Alloiooccus otitidis polynucleotides and polypeptides, useful for
PT treating and diagnosing diseases, drug screening assays and monitoring of
PT effects during drug clinical trials.

KW SAPHO syndrome; synovitis; acne; pustulosis; hypertosis; osteomyelitis;
KW uveitis; endophthalmitis; bone; joint; central nervous system; ELISA;
KW inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay;
KW dermatological; osteopathic; neuroprotectant.
XX
XX Propionibacterium acnes.
OS
PN WO200181581-A2.
XX
XX 01-NOV-2001.
PD
XX
PF 20-APR-2001; 2001WO-US012865.
XX
XX 21-APR-2000; 2000US-0199047P.
PR
PR 02-JUN-2000; 2000US-0208841P.
PR
PR 07-JUL-2000; 2000US-0216747P.
XX
XX (CORI-) CORIXA CORP.
PA
XX Skeiky YAW, Persing DH, Mitcham JL, Wang SS, Bhatia A;
PI L'maisonneuve J, Zhang Y, Jen S, Carter D;
XX
XX WPI; 2001-616774/71.
DR
DR N-PSDB; AAS59527.
XX
XX Propionibacterium acnes polypeptides and nucleic acids useful for
PT vaccinating against and diagnosing infections, especially useful for
PT treating acne vulgaris.
XX
XX Claim 3; SEQ ID NO 7017; 1069pp; English.
XX
XX Sequences AAU39105-AAU68017 represent Propionibacterium acnes immunogenic
CC polypeptides. The proteins and their associated DNA sequences are used in
CC the treatment, prevention and diagnosis of medical conditions caused by
CC P. acnes. The disorders include SAPHO syndrome (synovitis, acne,
CC pustulosis, hypertosis and osteomyelitis), uveitis and endophthalmitis.
CC P. acnes is also involved in infections of bone, joints and the central
CC nervous system, however it is particularly involved in the inflammatory
CC lesions associated with acne vulgaris. A method for detecting the
CC presence or absence of P. acnes in a patient comprises contacting a
CC sample with a binding agent that binds to the proteins of the invention
CC and determining the amount of bound protein in the sample. The
CC polypeptides may be used as antigens in the production of antibodies
CC specific for P. acnes proteins. These antibodies can be used to
CC downregulate expression and activity of P. acnes polypeptides and
CC therefore treat P. acnes infections. The antibodies may also be used as
CC diagnostic agents for determining P. acnes presence, for example, by
CC enzyme linked immunosorbent assay (ELISA). Note: The sequence data for
CC this patent did not form part of the printed specification, but was
CC obtained in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 277 AA;
Query Match 26.6%; Score 338.5; DB 4; Length 277;
Best Local Similarity 37.5%; Pred. No. 5.7e-29;
Matches 94; Conservative 42; Mismatches 100; Indels 15; Gaps 6;
QY 6 VRADELVFLQIAESRRCQAKELIAGKVTLTNNSTTFLRLEKPGHKYPLESICSLIGVE 65
DB 33 VRLDQALVEAGLARSRTLATRLVREGVTVNRVVVT-----KPAKVSTADVLV-VADEE 85
QY 66 RFSVRGAYKLTALDFFKIDVKSCICLDAGASTGGTDCILQHGASKVYAIADVKGQOLHE 125
DB 86 VWWSRGAKHLLGALDILGVQVPSRV-LDAGASTGGTGVVLSRGAKLVHADVGHQDQSP 144
QY 126 KLYTNEQVINIEGVNLR-----TASKOLIPPEVDILTIDVSFISLTILPSCIRWLKASG 180
DB 145 VLNRDPRVIVHEGLNRLDQPADLAVDGVWEPFDLVVGVDSVFSILTILPMSAVVSPDG 204
QY 181 IITIALIKPOPELYPDKI-KKGVVKETSIOYEAVEKIIHFCQSEGLGLIFIGVVPVVIKGP 239
DB 205 LMLLLVRFQFEVGRKALGAHGVVTDPLRLQIAIGVM-AAAVDLGWRMRDECDSPILPQD 263

QY 240 GNOEYLIYLKX 250
DB 264 GNVHEFFLLER 274
RESULT 24
ABM42341
ID ABM42341 standard; protein; 277 AA.
XX
XX AC ABM42341;
XX
XX 20-OCT-2003 (first entry)
XX
XX Propionibacterium acnes haemolysin-related polypeptide #7017.
XX
XX Acne vulgaris; antiseborrheic; dermatological; antibacterial;
KW immunostimulant; immune response; vaccine.
XX
XX Propionibacterium acnes.
OS
XX
XX WO2003033515-A1.
XX
XX 24-APR-2003.
XX
XX 11-OCT-2002; 2002WO-US032727.
PF
XX
XX 15-OCT-2001; 2001US-00978825.
PR
XX
XX (CORI-) CORIXA CORP.
PA
XX
XX Mitcham JL, Skeiky YAW, Persing DH, Bhatia A, Maisonneuve JL;
PI Zhang Y, Wang S, Jen S, Lodes MJ, Benson DR, Jones R, Carter D;
PI Barth B, Valliave-Douglas J;
XX
XX WPI; 2003-381789/36.
DR
DR N-PSDB; ACF64456.
XX
XX New Propionibacterium acnes polypeptides and polynucleotides encoding the
PT polypeptide, useful for diagnosing, preventing or treating acne vulgaris,
PT or for stimulating an immune response specific for a P. acnes protein.
XX
XX Claim 31; SEQ ID NO 7017; 1481pp; English.
XX
XX The invention relates to an isolated polynucleotide (ACF64435-ACF64733)
CC encoding a Propionibacterium acnes protein. The invention also relates to
CC polypeptides encoded by the polynucleotides (ABM35624-ABM64536) and to
CC immunogenic fragments of P. acnes polypeptides. The invention
CC additionally encompasses expression vectors and host cells comprising a
CC polynucleotide of the invention; antibodies against polypeptides of the
CC invention; fusion proteins comprising a polypeptide of the invention; a
CC method for stimulating an immune response specific for a P. acnes
CC polypeptide and an isolated T cell population comprising T cells prepared
CC via this method; a vaccine composition (comprising P. acnes polypeptides,
CC polynucleotides, antibodies, fusion proteins, T cell populations, or
CC antigen-presenting cells that express the polypeptide); a method and kit
CC for detecting or determining the presence or absence of P. acnes in a
CC patient; and a method for inhibiting the development of P. acnes in a
CC patient. The P. acnes polypeptides, polynucleotides, antibodies, fusion
CC proteins, T cell populations or antigen-presenting cells that express the
CC polypeptides are useful for diagnosing, preventing or treating acne
CC vulgaris, or for stimulating an immune response specific for a P. acnes
CC protein. The polynucleotides can also be used as probes or primers for
CC nucleic acid hybridisation. The vaccine composition is useful for the
CC stimulation of an immune response against P. acnes, or for treating acne,
CC and the kit is useful for performing a diagnostic assay. The present
CC sequence represents a specifically claimed polypeptide which is predicted
CC to be encoded by an ORF (open reading frame) contained within the P.
CC acnes polynucleotides of the invention. Note: The sequence data for this
CC patent did not form part of the printed specification, but was obtained
CC in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX


```
SQ      Sequence 277 AA;
Query Match      26.6%; Score 338.5; DB 6; Length 277;
Best Local Similarity 37.5%; Pred. No. 5.7e-29;
Matches 94; Conservative 42; Mismatches 100; Indels 15; Gaps 6;
QY      6 VRADLVFLQGLAESRQAKRLIMAGKVTLTNNSTPIRLKPGHKYPLESICSLIGVE 65
DB      33 VRLDQALVEAGLARSRLATRLVREGRTVNRVVT-----KPAKYSTADVLA-VADEE 85
QY      66 RVSRGAYKLLTALDFKIDVKSICLDAGASTGGFTDCLLOHGASKVVAIDVKGQOLHE 125
DB      86 VVWSRGAKHLGALDILGVQPSRV-LDAGASTGGFTQVLSRGAKLHVAVDVGHQDQSP 144
QY      126 KLYTNEQVINTEGVNLR-----TASKDLIPEVEDILTIDVSPISLTLLPSCIRWLKASG 180
DB      145 VLRNDPRVIVHEGLNRLDQADLAVDGVPEVDLVVGVGVDFISLTMLEPMSAVVSPDG 204
QY      181 IIALIKPQFELYPDKI-KKGVVKTSLQYEAKEKIIHFCQSELGLIFIGVVPSVIKGP 239
DB      205 LMLLVKPFQFVGRKALGAHGVWTDPALRLQAIAGWM-AAAVDLGWRMRDECDSPLPQD 263
QY      240 GNQEVLYLVKK 250
DB      264 GNVEHFFLLER 274

RESULT 25
ABU01632
ID      ABU01632 standard; protein; 198 AA.
XX
AC      ABU01632;
XX
DT      23-OCT-2003 (revised)
DT      11-FEB-2003 (first entry)
XX
DE      S. pneumoniae type 4 strain protein from coding region #1208.
XX
KW      Bacterial meningitis; pneumonia; sepsis; otitis media; ear infection;
KW      antiinflammatory; antibacterial; immunostimulant; auditory; respiratory;
KW      gene therapy; vaccine.
XX
OS      Streptococcus pneumoniae; type 4 strain.
XX
FN      WO200277021-A2.
XX
PD      03-OCT-2002.
XX
PF      27-MAR-2002; 2002WO-IB002163.
XX
PR      27-MAR-2001; 2001GB-00007658.
XX
PA      (CHIR-) CHIRON SPA.
PA      (GENO-) INST GENOMIC RES.
XX
PI      Massignani V, Tettelin H, Fraser C;
XX
DR      WPI; 2003-040579/03.
DR      N-PSDB; ABX06920.
XX
PT      New proteins and nucleic acid molecules from Streptococcus pneumoniae,
PT      useful as medicaments for treating or preventing a disease or infection
PT      due to streptococcus bacteria, such as pneumonia, sepsis, otitis media or
PT      ear infection.
XX
PS      Claim 1; SEQ ID NO 2416; 56pp; English.
XX
CC      The invention relates to a protein comprising or having at least 50%
CC      identity to any of the 2469 amino acid sequences, identified in the
CC      specification (available on a computer readable format), or its fragment,
CC      expressed from 2469 of 2489 identified DNA coding regions from the
CC      Streptococcus pneumoniae type 4 strain genomic sequence appearing as
CC      ABS56454. Also included are an antibody which binds one of the proteins,
CC      ABS56454.
```

```
CC      treating a patient by administering the protein, DNA or antibody (in a
CC      composition), a kit comprising first and second primers, which are the
CC      nucleic acid cited above or fragments between nucleotides 8-100 of a
CC      sequence not defined in the specification, for amplifying a target
CC      sequence contained within a Streptococcus nucleic acid sequence, where
CC      the first primer is substantially complementary to the target sequence
CC      and the second primer is substantially complementary to the complement of
CC      the target sequence, and where the parts of the primers having
CC      substantial complementarity define the termini of the target sequence to
CC      be amplified, assay comprising contacting a test compound with the
CC      protein, and determining whether the test compound binds to the protein
CC      and a Streptococcus pneumoniae bacterium, where one or more genes
CC      encoding the proteins has been rendered inactive. The proteins, nucleic
CC      acid molecules, antibody and compositions are useful as medicaments for
CC      treating or preventing a disease or infection due to streptococcus
CC      bacteria, particularly S. pneumoniae, such as pneumonia, sepsis, otitis
CC      media or ear infection. They are also useful in developing vaccines,
CC      diagnostics and antibiotics. The methods are useful for identifying
CC      immunodominant proteins. The present sequence is one of the 2469 proteins
CC      expressed by the identified coding regions from the genomic sequence.
CC      Note: The sequence data for this patent did not form part of the printed
CC      specification, but was obtained in electronic format directly from WIFO
CC      at ftp.wipo.int/pub/published_pct_sequences. (Updated on 23-OCT-2003 to
CC      standardise OS field)
XX
SQ      Sequence 198 AA;
Query Match      24.9%; Score 316; DB 6; Length 198;
Best Local Similarity 41.0%; Pred. No. 1.2e-26;
Matches 71; Conservative 33; Mismatches 67; Indels 2; Gaps 2;
QY      79 LDFEKIDVKSCICLDAGASTGGFTDCLLOHGASKVVAIDVKGQOLHEKLYTNEQV 138
DB      1 MQVFDLSVDGATTIDIGASTGGFTDVMQNSAKLVFAVDVGTNQAWKLQDPRVSMEQ 60
QY      139 VNLRTASKDLIPEVEDILTIDVSPISLTLLPSCIRWLKASGIIALIKPQFELYPDKI- 197
DB      61 ENFRYAEKTFEQEPSFASIDVSPISLTLLPALHRLVADQGVVALVKPQFAGREGI 120
QY      198 KGVVKTSLQYEAKEKIIHFCQSELGLIFIGVVPSVIKPGKQNOEYLVLYKK 250
DB      121 KNGIIRDAKVHQNVLSEVTAMA-VEVGFSVLGLDFSPIQGGHGNIEFLAYLKK 172

RESULT 26
AAW60947
ID      AAW60947 standard; protein; 191 AA.
XX
AC      AAW60947;
XX
DT      13-OCT-1998 (first entry)
XX
DE      Streptococcus pneumoniae encoded polypeptide.
XX
KW      coding region; ORF; open reading frame; antibacterial; infection;
KW      prevention; meningitis.
XX
OS      Streptococcus pneumoniae.
XX
FN      WO9819689-A1.
XX
PD      14-MAY-1998.
XX
PF      27-OCT-1997; 97WO-US019226.
XX
PR      01-NOV-1996; 96US-0029930P.
XX
PA      (SMIK ) SMITHKLINE BEECHAM CORP.
PA      (SMIK ) SMITHKLINE BEECHAM PLC.
XX
PI      Black MT, Hodgson JE, Knowles DJC, Lonetto MA, Nicholas RO;
PI      Reid RH, Zarfos FN;
XX
```

```

DR WPI; 1998-286586/25.
DR N-PSDB; AAV37347.
XX
PT New isolated nucleic acids from Streptococcus pneumoniae - useful, e.g.
PT for identifying anti-bacterial(s) for treatment and prevention of
PT meningitis.
XX
PS Claim 11; Page 50; 130pp; English.
XX
CC The sequence is that of the polypeptide encoded by a region isolated from
CC S. pneumoniae which shows homology to cytotoxin/hemolysin ORF2 tly. The
CC protein, or agonists of it, may be useful as an antibacterial for
CC treatment or prevention of infection, specifically caused by S.
CC pneumoniae (particularly meningitis) but possibly also Helicobacter
CC pylori (ulcers and gastric cancer). It may be of particular use before
CC insertion of an in-dwelling device or any other invasive procedure. The
CC protein, or nucleic acid encoding it, can also be used in vaccines to
CC induce a cellular and/or humoral immune response, or to screen for other
CC antibacterials. The DNA may also contain flanking sequences that are
CC potential sources of control elements for bacterial gene expression.
CC Detecting a sequence encoding the protein can be used diagnostically.
CC e.g. to detect a mutation for serotyping or classifying infectious agents
XX
XX Sequence 191 AA;
SQ
Query Match 24.5%; Score 311; DB 2; Length 191;
Best Local Similarity 42.2%; Pred. No. 4e-26;
Matches 70; Conservative 31; Mismatches 63; Indels 2; Gaps 2;
QY 86 VKSCICLDAGAGTGFTDCLLQHGASKVYVIDVKGQLEKLYTNEQVINEGVNRTAS 145
Db 1 VDGATTIDICAGTGFTDVMQLNSAKLVFAVDVGTNQLAWKLRQDPVVSMEQFNRYAE 60
QY 146 KDLPEVDILTIDVVSFISLTILPSCIRMLKASGIIILIKQFELYPDKI-KKGVVKE 204
Db 61 KTFDEQEPSASIDVVSFISLSLILPALHRLVADQGVVALVQFEGAGREIQGNGLIRD 120
QY 205 TSLQYEAVERKIIHFCQSELGLIFGVVPSVIKPGKNGQYLIYLKK 250
Db 121 AKIHQNVLESVTANA-VEAGFSVLGLDFSPIQGGHGNIELVLVLYKK 165
RESULT 27
AAW20115
ID AAW20115 standard; protein; 235 AA.
XX
AC AAW20115;
XX
DT 08-JUL-1997 (first entry)
XX
DE H. pylori secreted or periplasmic protein, 12505125.aa.
XX
KW Cytoplasmic; vaccine; prevention; treatment; infection; envelope;
KW identification; binding compound; bacterium; life cycle; activator;
KW bacteria; inhibitor; duodenal ulcer disease; chronic gastritis;
KW diagnosis.
XX
OS Helicobacter pylori.
XX
FN WO9640893-A1.
XX
PD 19-DEC-1996.
XX
PF 06-JUN-1996; 96WO-US009122.
XX
PR 07-JUN-1995; 95US-00487032.
XX
PR 01-APR-1996; 96US-00630405.
XX
PA (ASTR ) ASTRA AB.
XX
PI Smith D, Berglindh OT, Mellgaard BL;
XX
XX WPI; 1997-052306/05.
DR

```

```

DR N-PSDB; AAT67360.
XX
PT Helicobacter pylori nucleic acid sequences and related polypeptide(s) -
PT useful for vaccines to treat or prevent H. pylori infection, and to
PT detect Helicobacter.
XX
PS Claim 72; Page 347; 1481pp; English.
XX
CC This sequence represents a H. pylori secreted or periplasmic protein. The
CC protein may be used in a vaccine to prevent or treat H. pylori infection
CC or to identify H. pylori polypeptide-binding compounds, useful as
CC potential H. pylori life cycle activators or inhibitors. The genomic
CC sequence of H. pylori (ATCC 55679) was determined from overlapping
CC contigs generated by mechanically shearing the bacterial DNA. The
CC sequences were analysed for ORF of at least 180 nucleotides, and the
CC predicted coding regions defined by computer evaluation. To identify
CC likely H. pylori antigens for vaccine development, the amino acid
CC sequences predicted from various ORF were analysed for significant
CC homology to other known or exported membrane proteins. Having identified
CC and determined the sequences of interest, particular regions can be
CC isolated from H. pylori by PCR amplification for recombinant polypeptide
CC production, e.g. in E. coli hosts
XX
XX Sequence 235 AA;
SQ
Query Match 24.2%; Score 307.5; DB 2; Length 235;
Best Local Similarity 35.2%; Pred. No. 1.3e-25;
Matches 87; Conservative 40; Mismatches 105; Indels 15; Gaps 5;
QY 6 VRADELVFLQGLAESQAKRLIMAGKVTLTNNSTTILRLKPGHKYPLESICSLIGVE 65
Db 1 MRLDYALFNQHLANSSEKALVLRKQVLVKNVWVKSPFIVKEGDO-----IELIAPN 54
QY 66 RFVSRGAYKLLTALDFEKIDVKSICLDAGAGTGFTDCLLQHGASKVYVIDVKGQLEH 125
Db 55 LFVSRAGEKLGAFLEHDFIDFEKVVLDVGSKGFQVALLKGAKVLCVDVKGQMLE 114
QY 126 KLYTNEQVINEGVNRTASKOLIPREVDILTIDVVSFISLTILPSCIRMLKASGIIIL 185
Db 115 SLKNDQRIECYERCDINGFK--TPBKIDLALCDVVSFISLYCILALF--LPLSGEPLTL 169
QY 186 IKQFEL--YPDKIKKGVVKETSLOVEAVEKIIHFCQSELGLIFGVVPSVIKPGKNGQE 243
Db 169 FRPQFEVGTIKRKNKGVVMDKEALNALNFNHLKTK-DFQILITQESLVKGNNGVE 227
QY 244 YLIYLKK 250
Db 228 FFIHFEXR 234
RESULT 28
AAW20654
ID AAW20654 standard; protein; 246 AA.
XX
AC AAW20654;
XX
DT 14-JUL-1997 (first entry)
XX
DE H. pylori secreted or periplasmic protein, 02ge20116orf20.
XX
KW Cytoplasmic; vaccine; prevention; treatment; infection; identification;
KW binding compound; bacterium; life cycle; activator; bacteria; inhibitor;
KW duodenal ulcer disease; chronic gastritis; diagnosis; envelope.
XX
OS Helicobacter pylori.
XX
FN WO9640893-A1.
XX
PD 19-DEC-1996.
XX
PF 06-JUN-1996; 96WO-US009122.
XX
PR 07-JUN-1995; 95US-00487032.

```

```
PR 01-APR-1996; 96US-00630405.
XX (ASTR ) ASTRA AB.
XX Smith D, Berglindh OT, Mellgaard BL;
XX WPI; 1997-052306/05.
DR N-PSDB; AAT67907.
XX
PT Helicobacter pylori nucleic acid sequences and related polypeptide(s) -
ET useful for vaccines to treat or prevent H. pylori infection, and to
PT detect Helicobacter.
XX
PS Claim 72; Page 1077; 1481pp; English.
XX
CC The present sequence is a H. pylori secreted or periplasmic protein. The
CC protein may be used in a vaccine to prevent or treat H. pylori infection
CC or to identify H. pylori polypeptide binding compounds, useful as
CC potential H. pylori life cycle activators or inhibitors. The genomic
CC sequence of H. pylori (ATCC 55679) was determined from overlapping
CC contigs generated by mechanically shearing the bacterial DNA. The
CC sequences were analysed for ORF of at least 180 nucleotides, and the
CC predicted coding regions defined by computer evaluation. To identify
CC likely H. pylori antigens for vaccine development, the amino acid
CC sequences predicted from various ORF were analysed for significant
CC homology to other known or exported membrane proteins. Having identified
CC and determined the sequences of interest, particular regions can be
CC isolated from H. pylori by PCR amplification for recombinant polypeptide
CC production, e.g. in E. coli hosts
XX
SQ Sequence 246 AA;
Query Match 24.2%; Score 307.5; DB 2; Length 246;
Best Local Similarity 35.2%; Pred. No. 1.4e-25;
Matches 87; Conservative 40; Mismatches 105; Indels 15; Gaps 5;
QY 6 VRADELVFLQGLAESREQAERLIMAGKVTLTNNSTTIPRLRLEKPGHKYPLESICSLIGVE 65
Db 12 MRLDYALFNQHLANSREKAKALVLRKQVLRNKMVSKPSFIVKEDQ-----IELIAPN 65
QY 66 RFVSRGAYKLLTALDFFKIDVKSCICLDAGASTGGTDCLLQHGASKVYAIIDVGKQLHE 125
Db 66 LFVSRAGEKLGAFLETHFDFKFKWVLDVGASKGFSQVALLKGAKKVLCVDVGKQMLDE 125
QY 126 KLYTNEQVINIEGVNLTASKDLIPERVDILTIDVFSISLTILPSCIRMLKASGIILAL 185
Db 126 SLKNDQRIEYECEDIRGFK---TPEKIDALCDVFSISLYCILEAI---LPLSGEFTL 179
QY 186 IKPQFEL--YPDTKKGWVKETSQYEAWEKIHFQCSSELGLIFIGVVPVSKPGKNGQE 243
Db 180 FKQFQEVGRTIKRKKGVMDKEAILNALENFKNHLTK-DFQILTITQESLVKKGNGVE 238
QY 244 YLIYLKX 250
Db 239 FFIHFKR 245
RESULT 29
AAW98690
ID AAW98690 standard; protein; 235 AA.
XX
AC AAW98690;
XX
DT 31-MAR-1999 (first entry)
XX
DE H. pylori GHPO 611 protein.
XX
KW GHPO protein; Helicobacter infection; gastroduodenal disease; gastritis;
KW peptic ulcer disease.
XX
OS Helicobacter pylori.
XX
FN WO9843478-A1.
```

```
XX 08-OCT-1998.
XX
PF 01-APR-1998; 98WO-US006371.
XX
PR 01-APR-1997; 97US-00833457.
PR 24-JUN-1997; 97US-00881227.
PR 29-JUL-1997; 97US-00902615.
XX
PA (INMR ) MERIEUX ORAVAX PASTEUR MERIEUX SERUMS.
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Kleanthous H, Al-Garawi A, Miller C, Tomb J, Oomen RP;
XX
DR WPI; 1998-542293/46.
DR N-PSDB; AAX14409.
XX
PT New isolated Helicobacter polynucleotides - used to develop products for
PT the diagnosis, prevention and treatment of Helicobacter infections and
PT gastrointestinal diseases.
XX
PS Claim 8; Page 1433-1434; 2054pp; English.
XX
CC This sequence represents a Helicobacter pylori GHPO protein of the
CC invention. The polypeptides can be used for preventing or treating
CC Helicobacter infections, and gastroduodenal diseases associated with
CC these infections, including acute, chronic, and atrophic gastritis, and
CC peptic ulcer diseases, e.g. gastric and duodenal ulcers. They can also be
CC used for the production of antibodies. The products can also be used for
CC detection and diagnosis
XX
SQ Sequence 235 AA;
Query Match 22.9%; Score 290.5; DB 2; Length 235;
Best Local Similarity 33.6%; Pred. No. 1.1e-23;
Matches 83; Conservative 41; Mismatches 108; Indels 15; Gaps 5;
QY 6 VRADELVFLQGLAESREQAERLIMAGKVTLTNNSTTIPRLRLEKPGHKYPLESICSLIGVE 65
Db 1 MRLDYALFNSHLVNSREKAKALVLRKQVLRNKMVSKPSFIVKENDK-----IELIAEK 54
QY 66 RFVSRGAYKLLTALDFFKIDVKSCICLDAGASTGGTDCLLQHGASKVYAIIDVGKQLHE 125
Db 55 LFVSRAGEKLGAFLETHFDFKFKWVLDVGASKGFSQVALLKGAKKVLCVDVGKQMLDE 114
QY 126 KLYTNEQVINIEGVNLTASKDLIPERVDILTIDVFSISLTILPSCIRMLKASGIILAL 185
Db 115 SLKQDKRIEYECEDIRGFK---TPETIDALCDVFSISLYILEAI---LPLSDEFLL 168
QY 186 IKPQFELYP--DKIKKGWVKETSQYEAWEKIHFQCSSELGLIFIGVVPVSKPGKNGQE 243
Db 169 FKQFQEVGRTIKRKKGVVDKEAILNALENFKNHLTK-DFQILKIQESLVKKGKNGVE 227
QY 244 YLIYLKX 250
Db 228 FFIHFKR 234
RESULT 30
ABP03533
ID ABP03533 standard; protein; 105 AA.
XX
AC ABP03533;
XX
DT 25-JUN-2002 (first entry)
XX
DE Human ORFX protein sequence SEQ ID NO:7048.
XX
KW Human; open reading frame; ORFX; gene therapy; cancer; cirrhosis;
KW hyperproliferative disorder; psoriasis; benign tumour; haemorrhage;
KW degenerative disorder; osteoarthritis; neurodegenerative disorder;
KW cardiovascular disease; diabetes mellitus; systemic lupus erythematosus;
KW hypertension; hypothyroidism; cholesterol ester storage disease;
```

immune deficiency; immune disorder; infectious disease;
KW autoimmune disorder; rheumatoid arthritis; autoimmune thyroiditis;
KW myasthenia gravis.
XX
XX Homo sapiens.
OS
XX WO200192523-A2.
PN
XX 06-DEC-2001.
XX
XX 29-MAY-2001; 2001WO-US010836.
XX
XX 30-MAY-2000; 2000US-0206132P.
PR
XX 29-AUG-2000; 2000US-0228716P.
XX
XX (CURA-) CURAGEN CORP.
PA
XX
XX Shimkets RA, Leach MD;
PI
XX WPI; 2002-106308/14.
DR
XX N-PSDB; ABN19285.
XX
XX Novel human polypeptides and polynucleotides useful for diagnosing,
PT preventing and treating cardiovascular disease, neurodegenerative,
PT hyperproliferative disorders and autoimmune disorders.
XX
XX Disclosure; SEQ ID NO 7048; 1037pp; English.
PS
XX
XX The present invention describes substantially purified human proteins
CC (referred to as open reading frame, ORFX, where X is 1-11491 (see Table 1
CC in the specification). ABN15762 to ABN27252 encode the human ORFX
CC proteins given in ABP00010 to ABP11500. ORFX proteins are useful for
CC treating or preventing a pathology associated with an ORFX-associated
CC disorder in humans, and in the manufacture of a medicament for treating a
CC syndrome associated with ORFX-associated disorder. ORFX polynucleotide
CC sequences can be used in gene therapy. ORFX sequences can be used in the
CC treatment of cancer, hyperproliferative disorders, cirrhosis of liver,
CC psoriasis, benign tumours, keloid, degenerative disorders, haemorrhage,
CC osteoarthritis, neurodegenerative disorders, disorders related to organ
CC transplantation, cardiovascular diseases, diabetes mellitus, systemic
CC lupus erythematosus, hypertension, hypothyroidism, cholesterol ester
CC storage disease, various immune deficiencies and disorders, infectious
CC diseases, autoimmune disorders such as multiple sclerosis, rheumatoid
CC arthritis, autoimmune thyroiditis, myasthenia gravis, graft-versus-host
CC disease and autoimmune inflammatory eye disease. ORFX proteins are also
CC useful for treating burns, incisions, ulcers, for treating osteoporosis,
CC bone degenerative disorders, or periodontal disease, and for gut
CC protection or regeneration and treatment of lung or liver fibrosis,
CC reperfusion injury in various tissues and conditions resulting from
CC systemic cytokine damage. N.B. The sequence data for this patent did not
CC form part of the printed specification, but was obtained in electronic
CC format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
XX Sequence 105 AA;
SQ

Query Match 11.2%; Score 142.5; DB 5; Length 105;
Best Local Similarity 41.5%; Pred. No. 1.3e-07;
Matches 39; Conservative 11; Mismatches 37; Indels 7; Gaps 2;

QY 6 VRADELVFLQGLAESREQAKRLIMAGKVTLTNNSTIPLRLKPGHKYPLESICSLIGVE 65
Db 19 VRLDQALVEAGLARSRTLRLVREGRTVNRVVVT-----KPAKVKSTTDVLXVADEE 72

QY 66 RYVSRGAYKLLTALDFFKIDVKSCICLDAGASTG 99
Db 73 VMTSRAAHKLLGALDILGVQVPSRV-LDAGASTG 105

RESULT 31
ABBS4612
ID ABBS4612 standard; protein; 256 AA.
XX
XX ABBS4612;

XX 29-AUG-2003 (revised)
DT 16-MAY-2002 (first entry)
XX
XX Lactococcus lactis protein rsuB.
DE
XX Biosynthesis; biodegradation; lactic bacterium; yogurt; cheese.
KW
XX Lactococcus lactis; IL1403.
OS
XX FR2807446-A1.
PN
XX 12-OCT-2001.
PD
XX 11-APR-2000; 2000FR-00004630.
PF
XX 11-APR-2000; 2000FR-00004630.
PR
XX (INRG) INRA INST NAT RECH AGRONOMIQUE.
PA
XX Bolotine A, Sorokine A, Renault P, Ehrlich SD;
PI
XX WPI; 2002-043418/06.
XX
DR New nucleotide sequence useful in the identification or Lactococcus
XX lactis and related species.
PT
XX Claim 6; SEQ ID NO 1314; 2504pp; French.
PS
XX The present invention is related to a Lactococcus lactis nucleotide
CC sequence (ABA90521) and related proteins (ABBS3300-ABBS5621). The nucleic
CC acid sequence is useful in the detection and/or amplification of nucleic
CC acid sequence, particularly to identify Lactococcus lactis or related
CC species. The proteins of the invention are useful for the biosynthesis or
CC biodegradation of a composition of interest. The invention helps research
CC in lactic bacteria, particularly useful in the production of yogurt and
CC cheese. Note: The sequence data for this patent is based on equivalent
CC patent WO200177334 (published 19-OCT-2001) which is available in
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences. (Updated on 29-AUG-2003 to
CC standardise OS field)
XX
XX Sequence 256 AA;
SQ

Query Match 8.4%; Score 106.5; DB 5; Length 256;
Best Local Similarity 28.4%; Pred. No. 0.0053;
Matches 55; Conservative 29; Mismatches 67; Indels 43; Gaps 11;

QY 5 KVRADLVFLQGLAESREQAKRLIMAGKVTLTNNSTIPLRLKPGHKYPLESICSLIGV 64
Db 18 KMRINKYLAHAGVA-SRRKAELILAGKVTNN-----VP--MTNLGYQVSTGDIVVNGV 70

QY 65 ERVSRGAYKLL-----TALDFFKIDVKSCIC-----LDAGAS----- 97
Db 71 AVTNEEPVYLLNKPGRYISSVSDDKGRTQVMDLMP-QVKERYFPVGRLDWDTSGLLLT 129

QY 98 -TGGFTDCLL--QHGAASKYVAIDVGKQLH-EKLYTNEQVINTEGVNLRTASKDLIPEEV 153
Db 130 NDGEFTNMVTHPRHGVDKVVAKV-EGQANKENLRPLTGLMTTEGKKVSPARVEIIKQE- 187

QY 154 DILTIDVSISLT 167
Db 188 --KTKNHSIVSLTI 199

RESULT 32
ABP79451
ID ABP79451 standard; protein; 627 AA.
XX
XX ABP79451;
XX
XX 07-MAR-2003 (first entry)
XX

DE N. gonorrhoeae amino acid sequence SEQ ID 5432.
XX Antibacterial; infection; vaccine; gene therapy.
XX Neisseria gonorrhoeae.
XX WO200279243-A2.
XX 10-OCT-2002.
XX
XX 12-FEB-2002; 2002WO-IB002069.
XX
XX 12-FEB-2001; 2001GB-00003424.
XX (CHIR-) CHIRON SPA.
XX
XX Fontana MR, Pizza M, Massignani V, Monaci E;
XX WPI; 2003-058415/05.
XX N-PSDB; ABZ40421.
XX
XX New protein from Neisseria gonorrhoeae, useful for the manufacture of a
XX medicament for treating or preventing N. gonorrhoeae infection.
XX
XX Disclosure; Page 584-585; 815pp; English.
XX
XX The present invention relates to proteins from Neisseria gonorrhoeae.
XX Also disclosed are the nucleic acid molecules encoding the proteins and
XX antibodies that specifically bind to the proteins. The composition
XX comprising the protein, nucleic acid or antibody is useful for the
XX manufacture of a medicament for treating or preventing N. gonorrhoeae
XX infection, this may be in the form of a vaccine or gene therapy.
XX Sequences given in records ABP76736-ABP81046 represent nucleic acid
XX molecules of the invention
XX
SQ Sequence 627 AA;

Query Match 7.4%; Score 93.5; DB 6; Length 627;
Best Local Similarity 23.7%; Pred. No. 0.57;
Matches 62; Conservative 32; Mismatches 89; Indels 79; Gaps 12;

QY 10 ELVFLQGLAESRQAKRLIMAGKVTLTN-----NSTTIPRLKPGHKYPLESICSLIGV 64
Db 241 DVNALQRLKEAEAKAKIELSSGQQTINLPYITMDATGPKHLAMKTRAKFESL-----V 295

QY 65 ERFVSRGAYKLLTALDFPKIDVKSICLDAGASTGFTDCLLOHGASKVYVIDVGKQQLH 124
Db 296 EDLIARSIEPCRTALK-----DAGLSTGDDIDVILVGQGRMPKV----- 335

QY 125 EKYLTNEQVINIEGVNLTARTASKDLIPEV-----DILTIDVSFISLTL 167
Db 336 -----QEAQVDFGKEPR---KDVNPDEAVAGNAIQGEVLGSGRSVLLLDVTFPLSLGI 387

QY 168 ILPSCIRWLKASGIIIALIKPQELYPDKIKKGVKETSIOYRAVEKIHFCQSE----- 222
Db 388 -----ETMGGVMTKLIQKNTTI-PTKASQ-VFSTAEDNQSAV--TIHVLQGERBRAS 435

QY 223 ----LGLIFGVPSVIKPGK 240
Db 436 ANKSLGQFNLG---DIAPAPRG 454

RESULT 33
ABU37433
ID ABU37433 standard; protein; 642 AA.
XX
XX AC ABU37433;
XX
XX 23-OCT-2003 (revised)
DT 19-JUN-2003 (first entry)
XX
XX Protein encoded by Prokaryotic essential gene #22960.
DE
XX

XX Antisense; prokaryotic essential gene; cell proliferation; drug design.
XX Neisseria gonorrhoeae.
XX WO200277183-A2.
XX 03-OCT-2002.
XX
XX 21-MAR-2002; 2002WO-US009107.
XX
XX 21-MAR-2001; 2001US-00815242.
XX 06-SEP-2001; 2001US-00948993.
XX 25-OCT-2001; 2001US-0342923P.
XX 08-FEB-2002; 2002US-00072851.
XX 06-MAR-2002; 2002US-0362699P.
XX
XX (ELIT-) ELITRA PHARM INC.
XX
XX Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;
XX Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;
XX WPI; 2003-029926/02.
XX N-PSDB; ACA41303.
XX
XX New antisense nucleic acids, useful for identifying proteins or screening
XX for homologous nucleic acids required for cellular proliferation to
XX isolate candidate molecules for rational drug discovery programs.
XX
XX Claim 25; SEQ ID NO 65357; 1766pp; English.
XX
XX The invention relates to an isolated nucleic acid comprising any one of
XX the 6213 antisense sequences given in the specification where expression
XX of the nucleic acid inhibits proliferation of a cell. Also included are:
XX (1) a vector comprising a promoter operably linked to the nucleic acid
XX encoding a polypeptide whose expression is inhibited by the antisense
XX nucleic acid; (2) a host cell containing the vector; (3) an isolated
XX antisense nucleic acid; (4) an antibody capable of specifically binding
XX the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
XX proliferation or the activity of a gene in an operon required for
XX proliferation; (7) identifying a compound that influences the activity of
XX the gene product or that has an activity against a biological pathway
XX required for proliferation, or that inhibits cellular proliferation; (8)
XX identifying a gene required for cellular proliferation or the biological
XX pathway in which a proliferation-required gene or its gene product lies
XX or a gene on which the test compound that inhibits proliferation of an
XX organism acts; (9) manufacturing an antibiotic; (10) profiling a
XX compound's activity; (11) a culture comprising strains in which the gene
XX product is overexpressed or underexpressed; (12) determining the extent
XX to which each of the strains is present in a culture or collection of
XX strains; or (13) identifying the target of a compound that inhibits the
XX proliferation of an organism. The antisense nucleic acids are useful for
XX identifying proteins or screening for homologous nucleic acids required
XX for cellular proliferation to isolate candidate molecules for rational
XX drug discovery programs, or for screening homologous nucleic acids
XX required for proliferation in cells other than S. aureus, S. typhimurium,
XX K. pneumoniae or P. aeruginosa. The present sequence is encoded by one of
XX the target prokaryotic essential genes. Note: The sequence data for this
XX patent did not form part of the printed specification, but was obtained
XX in electronic format directly from WIPO at
XX ftp.wipo.int/pub/published_pct_sequences. (Updated on 23-OCT-2003 to
XX standardise OS field)
XX
SQ Sequence 642 AA;

Query Match 7.4%; Score 93.5; DB 6; Length 642;
Best Local Similarity 23.7%; Pred. No. 0.59;
Matches 62; Conservative 32; Mismatches 89; Indels 79; Gaps 12;

QY 10 ELVFLQGLAESRQAKRLIMAGKVTLTN-----NSTTIPRLKPGHKYPLESICSLIGV 64
Db 256 DVNALQRLKEAEAKAKIELSSGQQTINLPYITMDATGPKHLAMKTRAKFESL-----V 310

Qy	65	ERFVSRGAYKLLTALDFFPKIDVKSCICLDAGASTGGFTDCLLOHGASKVYADVGVQOLH	124
Db	311	EDLIARSIEPCTALK-----DAGLSTGTDIDDVILVGGOSRMPKV-----	350
Qy	125	EKLYTNEOVINIEGVNLTASKDLPIEV-----DILTIDVFSFISLT	167
Db	351	-----QEAVKOFFGKEPR---KDVNPDEAVAVGAAIQGEVLSGGRSVDILLDVTPLSLGI	402
Qy	168	ILPSCIRWLKASGIIIALIKPOFELYPPKIKKGVVKETSLQYAEVKEIHFQOSE	222
Db	403	-----ETWGGVTKLIQKNTTII-PIKASQ-VFSTAEDNQSAV--THVLOGERERAS	450
Qy	223	-----LGLIFIGVVPVSIKGPKG	240
Db	451	ANKSLGOFNLG---DIAPAPRG	469

RESULT 34

ABU37860	ABU37860 standard; protein; 642 AA.
ID	ABU37860
XX	
AC	ABU37860;

DT 19-JUN-2003 (first entry)

DE Protein encoded by Prokaryotic essential gene #233387.

AA Antisense; prokaryotic essential gene; cell proliferation; drug design.
 KW
 XX
 OS *Neisseria meningitidis*.

PN WO200277183-A2.

PD 03-OCT-2002.

PF 21-MAR-2002; 2002WO-US009107.

21-MAR-2001: 2001US-00815242.

09-SEP-2001; 2001US-00948993.
25-OCT-2001; 2001US-0342923P.

PK 08-FEB-2002; 2002US-00072851.
PR 06-MAR-2002; 2002US-0362699P

XX
PA (ELITT-) ELITRA PHARM INC

XX
pt Wang Y. Zamudio C. Maldonado C.

PL Wall D, Trawick JD, Car
yy

DR WPI; 2003-029926/02.

XX

PT for homologous nu

XX

XX

the 6213 antisense sequences given in the specification where expression of the nucleic acid inhibits proliferation of a cell. Also included are: (1) a vector comprising a promoter operably linked to the nucleic acid encoding a polypeptide whose expression is inhibited by the antisense nucleic acid; (2) a host cell containing the vector; (3) an isolated polypeptide or its fragment whose expression is inhibited by the antisense nucleic acid; (4) an antibody capable of specifically binding the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular proliferation or the activity of a gene in an operon required for proliferation; (7) identifying a compound that influences the activity of the gene product or that has an activity against a biological pathway required for proliferation, or that inhibits cellular proliferation; (8) identifying a gene required for cellular proliferation or the biological pathway in which a proliferation-required gene or its gene product lies or a gene on which the test compound that inhibits proliferation of an

organism acts; (9) manufacturing an antibiotic; (10) profiling a compound's activity; (11) a culture comprising strains in which the gene product is overexpressed or underexpressed; (12) determining the extent to which each of the strains is present in a culture or collection of strains; or (13) identifying the target of a compound that inhibits the proliferation of an organism. The antisense nucleic acids are useful for identifying proteins or screening for homologous nucleic acids required for cellular proliferation to isolate candidate molecules for rational drug discovery programs, or for screening homologous nucleic acids required for proliferation in cells other than *S. aureus*, *S. typhimurium*, *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of the target prokaryotic essential genes. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at www.int/pub/published/pct_sequences

Query Match	7.4%;	Score 93.5;	DB 6;	Length 642;
Best Local Similarity	23.7%;	Pred. NO. 0.59;		
Matches	62;	Conservative	32;	Mismatches 89;
				Indels 79;
				Gaps 12

QY 10 ELVFLQGLAESREQARLLIMAGKVTLTN-----NSTTIPURLEKPGHKYPJESICSIGV 64
:: :: :: :: ::
DB 256 DVMALORLKEAAEAKKIELSSGGOOTEINLPYTMDATGPXHLAMKTIRAKESL-----V 310

```
QY      65   ERFVSRGAYKLLTALDFFKIDVKSCICLDAGASTGGFTDCLLQHGAASKVYAIDVVGKGQLH 124
```

QY 125 EKLYTNEQVINTEGVNLRTASKDLIPEEV-----DILTIDVSFISLTLL 167

QY 168 ILPSCIRLWLKASGIIIALIKPQFELYDPDKIKGVVKETSLQYEAIVEKIITHECOSE----- 2222

Db 403 -----ETMGVMTKLIQKNTTI-

Qy 223 ---LGLIFIGVPSVIKPKG 240

RESULT 35

ABU25088
ID ABU25088 standard; protein; 615 AA.
XX
AC ABU25088;

DT 19-JUN-2003 (first entry)

DE Protein encoded by prokaryotic essential gene #10615.

Antisense: prokaryotic essential gene: cell proliferation: drug design.

clostridium difficile.

AA
PN
WQ200277183-A2XX
PD
03-OCT-2002XX
PF 21-MAR-2002 2002WO-IIS009107XX
PR 21-MAR-2001. 2001US-00815242

PR 06-SEP-2001; 2001US-00948993.
PR 25-OCT-2001; 2001US-03428332.

PR 08-FEB-2002; 2002US-00072851.
PP 06 MAR 2002; 2002US 0362588P

[illegible]

1000

PI Wall D, Trawick JD, Carr

Db 330 -----PCSGFVLRXP-----DIKYATEKDIIHKL-AE 357

QY 223 LGLIFIGVVPVSIK 236

Db 358 IQLAILDDVSQLVK 371

RESULT 37

ABU32431

ID ABU32431 standard; protein; 444 AA.

XX ABU32431;

XX 19-JUN-2003 (first entry)

XX Protein encoded by Prokaryotic essential gene #17958.

DE Antisense; prokaryotic essential gene; cell proliferation; drug design.

XX

XX Listeria monocytogenes.

OS

XX WO200277183-A2.

XX

XX 03-OCT-2002.

XX

XX 21-MAR-2002; 2002WO-US009107.

XX

XX 21-MAR-2001; 2001US-00815242.

PR

XX 06-SEP-2001; 2001US-00948993.

PR

XX 25-OCT-2001; 2001US-0342923P.

PR

XX 08-FEB-2002; 2002US-00072851.

PR

XX 06-MAR-2002; 2002US-0362699P.

XX

XX (ELIT-) ELITRA PHARM INC.

PA

XX Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;

PI

XX Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;

XX

XX WPI; 2003-029926/02.

DR

XX N-PSDB; ACA36301.

XX

XX New antisense nucleic acids, useful for identifying proteins or screening

PT

PT for homologous nucleic acids required for cellular proliferation to

PT

PT isolate candidate molecules for rational drug discovery programs.

XX

XX Claim 25; SEQ ID NO 60355; 1766pp; English.

XX

XX The invention relates to an isolated nucleic acid comprising any one of

CC

CC the 6213 antisense sequences given in the specification where expression

CC

CC of the nucleic acid inhibits proliferation of a cell. Also included are:

CC

CC (1) a vector comprising a promoter operably linked to the nucleic acid

CC

CC encoding a polypeptide whose expression is inhibited by the antisense

CC

CC nucleic acid; (2) a host cell containing the vector; (3) an isolated

CC

CC polypeptide or its fragment whose expression is inhibited by the

CC

CC antisense nucleic acid; (4) an antibody capable of specifically binding

CC

CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular

CC

CC proliferation or the activity of a gene in an operon required for

CC

CC proliferation; (7) identifying a compound that influences the activity of

CC

CC the gene product or that has an activity against a biological pathway

CC

CC required for proliferation, or that inhibits cellular proliferation; (8)

CC

CC identifying a gene required for cellular proliferation or the biological

CC

CC pathway in which a proliferation-required gene or its gene product lies

CC

CC or a gene on which the test compound that inhibits proliferation of an

CC

CC organism acts; (9) manufacturing an antibiotic; (10) profiling a

CC

CC compound's activity; (11) a culture comprising strains in which the gene

CC

CC product is overexpressed or underexpressed; (12) determining the extent

CC

CC to which each of the strains is present in a culture or collection of

CC

CC strains; or (13) identifying the target of a compound that inhibits the

CC

CC proliferation of an organism. The antisense nucleic acids are useful for

CC

CC identifying proteins or screening for homologous nucleic acids required

CC

CC for cellular proliferation to isolate candidate molecules for rational

CC

CC drug discovery programs, or for screening homologous nucleic acids

CC

CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,

CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of

CC the target prokaryotic essential genes. Note: The sequence data for this

CC patent did not form part of the printed specification, but was obtained

CC in electronic format directly from WIPO at

CC ftp.wipo.int/pub/published_pct_sequences

XX

SQ Sequence 444 AA;

Query Match 7.3%; Score 92.5; DB 6; Length 444;

Best Local Similarity 22.0%; Pred. No. 0.44;

Matches 56; Conservative 34; Mismatches 87; Indels 77; Gaps 9;

QY 16 GLAESRQAKRLIMAGKVTLTNNSTIP-----LRLEKPG- 50

Db 162 GIEQLREICLAFIVAPHOSIRVNOTEIPEQLIKELNDQGITVTRNEFIEALLVEKGV 221

QY 51 --HKYPLESICSLIGVERFVSRGAYKLLTALDFKIDVKSCICLDAGASTGGTDCILQ- 107

Db 222 AETKAYKDKGCSIQDESSMLAAYALQLEDNL-----TVLDACAAPGGKTHIAEK 271

QY 108 -HGASKVYADV--GKGQHEKLYTNEQVINIEGVNL--RTASKDLIPBEVDILTIDVSF 162

Db 272 MHGTGMVHALDIHEKTKLIDEAAKQLQNLNIRTAHLDAARTASTWPEPETEDRILVDA-- 329

QY 163 ISLTILPSCIRMLKASGIIIALIKPQFELYDPKIKKGVVKETSLOYEAVEKIIHPCQSE 222

Db 330 -----PCSGFVLRXP-----DIKYATEKDIIHKL-AE 357

QY 223 LGLIFIGVVPVSIK 236

Db 358 IQLAILDDVSQLVK 371

RESULT 38

ABE61456

ID ABE61456 standard; protein; 530 AA.

XX

AC ABE61456;

XX

XX 26-MAR-2002 (first entry)

XX

XX *Drosophila melanogaster* polypeptide SEQ ID NO 11160.

DE

XX *Drosophila*; developmental biology; cell signalling; insecticide;

XX

XX pharmaceutical.

XX

OS *Drosophila melanogaster*.

XX

XX WO200171042-A2.

XX

XX 27-SEP-2001.

XX

XX 23-MAR-2001; 2001WO-US009231.

XX

XX 23-MAR-2000; 2000US-0191637P.

PR

XX 11-JUL-2000; 2000US-00614150.

XX

XX (PEKE) PE CORP NY.

PA

XX Venter JC, Adams M, Li PWD, Myers EW;

PI

XX WPI; 2001-656860/75.

DR

XX N-PSDB; ABL05559.

XX

XX New isolated nucleic acid detection reagent for detecting 1000 or more

PT

PT genes from *Drosophila* and for elucidating cell signalling and cell-cell

PT

PT interactions.

XX

XX Disclosure; SEQ ID NO 11160; 21pp + Sequence Listing; English.

XX

XX The invention relates to an isolated nucleic acid detection reagent

CC

CC capable of detecting 1000 or more genes from *Drosophila*. The invention is

CC

CC useful in developmental biology and in elucidating cell signalling and

cell-cell interactions in higher eukaryotes for the development of insecticides, therapeutics and pharmaceutical drugs. The invention discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-ABB572072). The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences

CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 530 AA;

Query Match 7.1%; Score 90.5; DB 4; Length 530;
Best Local Similarity 26.0%; Pred. No. 0.97;
Matches 46; Conservative 29; Mismatches 63; Indels 39; Gaps 9;

Qy 84 IDVKSCICLDAGASTGGTDCCLLQHGKASKVAIDVGK-GQLHEKLYTNEQVIN----- 135
Db 176 VDFQDKIVLDVGAGSGILSPFAVQAGAKVYAIASNAWYAQAQQLVESNNVQHKISVIPG 235
Qy 136 -IEGVNLTASKDLIPREVDILTIDVSFISL--TLILPSCI--RWLKASGIIIIALIKPQ 189
Db 236 KIEEIEL-----PEKVDVIISEPMGYMLYNERMLETYLHARKWL-----KQP 277
Qy 190 FELYP--DKIKKGWVKETSLQYEAKEKLIHPCSELGLIFIGVPSVIKPKGNQY 244
Db 278 GKMYPTHGDLHIAPFSDSELYSEQYNKANFWYQS----AFHGVDLTTIH-KEGMKEY 329

RESULT 39
ABB58636
ID ABB58636 standard; protein; 1209 AA.
XX
AC ABB58636;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster polypeptide SEQ ID NO 2700.
XX
KW Drosophila; developmental biology; cell signalling; insecticide; pharmaceutical.
XX
OS Drosophila melanogaster.
XX
PN W0200171042-A2.
XX
PD 27-SEP-2001.
XX
PF 23-MAR-2001; 2001WO-US009231.
XX
PR 23-MAR-2000; 2000US-0191637P.
XX
PR 11-JUL-2000; 2000US-00614150.
XX
PA (PEKE) PE CORP NY.
PI Venter JC, Adams M, Li PWD, Myers EW;
XX
DR WPI; 2001-656860/75.
DR N-PSDB; ABL02739.
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more genes from Drosophila and for elucidating cell signalling and cell-cell interactions.
XX
PS Disclosure; SEQ ID NO 2700; 21pp + Sequence Listing; English.
XX
CC The invention relates to an isolated nucleic acid detection reagent capable of detecting 1000 or more genes from Drosophila. The invention is useful in developmental biology and in elucidating cell signalling and cell-cell interactions in higher eukaryotes for the development of insecticides, therapeutics and pharmaceutical drugs. The invention discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-ABB572072). The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly

from WIPO at ftp.wipo.int/pub/published_pct_sequences

CC
XX
SQ Sequence 1209 AA;

Query Match 7.1%; Score 90; DB 4; Length 1209;
Best Local Similarity 23.7%; Pred. No. 3.7;
Matches 45; Conservative 36; Mismatches 65; Indels 44; Gaps 10;

Qy 8 ADELVFLQGLAESREQAKRLIMAGKVTLTNNSTTIPRLKPGHKYPLESTCSLIGVERP 67
Db 834 ADRIINYENLHRTHPQ-----GSVPVDNGASL-----EHLLPQSRILVIGNREW 878
Qy 68 VSRGAYKLLTALDPFKIDVKSCICLDAGASTGGFTD--CLLQHGASKVYAI-DVKGKGLH 124
Db 879 MERNAIIV-----PLEISDCMTH-----ERKGTAVLCALNGQLVCMFAVSDMKPEAH 928
Qy 125 EKLYTNEQ-----VINIEGVNLTASKDLIPREVDILTIDVSFISLTILPS-----CIRWL 176
Db 929 LAVYTLKRMGIDVVLTLTGDNKNTAAS--IAREVGIRTVYAE-----VLPCHKVAKIQRI 980
Qy 177 KASGIIIALI 186
Db 981 QANGIRVAMV 990

RESULT 40
ADA35795
ID ADA35795 standard; protein; 304 AA.
XX
AC ADA35795;
XX
DT 20-NOV-2003 (first entry)
XX
DE Acinetobacter baumannii protein #2956.
XX
KW Acinetobacter baumannii; bacterial disease; antibacterial; vaccine; plant biocontrol agent.
XX
OS Acinetobacter baumannii.
XX
PN US6562958-B1.
XX
PD 13-MAY-2003.
XX
PF 04-JUN-1999; 99US-00328352.
XX
PR 09-JUN-1998; 98US-0088701P.
XX
PA (GENO-) GENOME THERAPEUTICS CORP.
PI Breton G, Bush D;
XX
DR WPI; 2003-576092/54.
DR N-PSDB; ADA31669.
XX
PT New Acinetobacter baumannii proteins and nucleic acids, useful as reagents for diagnosing a bacterial disease, as components of antibacterial vaccines, as targets for antibacterial drugs, or as biocontrol agents for plants.
XX
PS Example; SEQ ID NO 7082; 328pp; English.
XX
CC The invention relates to isolated Acinetobacter baumannii nucleic acids. The A. baumannii nucleic acids and polypeptides are useful as reagents for diagnosing a bacterial disease, as components of antibacterial vaccines, as targets for antibacterial drugs, to detect the presence of A. baumannii and other Acinetobacter species in a sample, in screening compounds for the ability to interfere with the A. baumannii life cycle or to inhibit A. baumannii infection, and as biocontrol agents for plants. The present sequence represents the amino acid sequence of an A. baumannii protein.
XX
SQ Sequence 304 AA;

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 7, 2004, 16:47:24 ; Search time 16.9435 Seconds
(without alignments)
833.792 Million cell updates/sec

Title: US-10-009-919A-1_COPY_1_50
Perfect score: 245
Sequence: 1 MAKHKVRADELVFQGLAES.....GKVTLTNNSTTIPRLLEKPG 50

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues
Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A Geneseq_29Jan04:*
1: Geneseqp1980s:*
2: Geneseqp1990s:*
3: Geneseqp2000s:*
4: Geneseqp2001s:*
5: Geneseqp2002s:*
6: Geneseqp2003as:*
7: Geneseqp2003bs:*
8: Geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	245	100.0	251	4 AAB48003	Aab48003 L. intrac
2	95	38.8	270	2 AAW55094	Aaw55094 Streptoco
3	95	38.8	270	5 ABP54588	Abp54588 S. pneumo
4	95	38.8	270	7 ADC45145	Adc45145 S. pneumo
5	93	38.0	275	5 ABP25556	Abp25556 Streptoco
6	90.5	36.9	227	7 ADC97152	Adc97152 E. faeciu
7	90	36.7	275	5 ABP25555	Abp25555 Streptoco
8	88	35.9	274	5 ABB48304	Abb48304 Listeria
9	87	35.5	191	6 ABU25523	Abu25523 Protein e
10	77	31.4	66	3 AAY77897	Aay77897 B. stearo
11	76	31.0	280	5 ABP65873	Abp65873 Bifidobac
12	75	30.6	139	6 ADB11952	Adb11952 Alloiococ
13	75	30.6	201	6 ADB11950	Adb11950 Alloiococ
14	75	30.2	209	5 ABG98423	Abg98423 Thermus t
15	74	30.2	137	2 AAW77671	Aaw77671 30S ribos
16	74	30.2	195	3 AAY70150	Aay70150 Staphyloc
17	74	30.2	195	5 AAE25699	Aae25699 Staphyloc
18	74	30.2	200	4 AAU37185	Aau37185 Staphyloc
19	74	30.2	200	4 AAU33908	Aau33908 Staphyloc
20	74	30.2	200	4 AAU36985	Aau36985 Staphyloc
21	74	30.2	200	5 AAO20048	Aao20048 Staphyloc
22	74	30.2	200	5 AAO20069	Aao20069 Staphyloc
23	74	30.2	200	6 ABU16110	Abu16110 Protein e
24	73	29.8	200	6 ABU17818	Abu17818 Protein e
25	73	29.8	200	6 ABU43088	Abu43088 Protein e

26	73	29.8	220	5 ABP40228	Abp40228 Staphyloc
27	71	29.0	201	4 AAU33534	Aau33534 Enterococ
28	71	29.0	203	4 AAU35286	Aau35286 Enterococ
29	71	29.0	203	6 ABU14521	Abu14521 Protein e
30	70	28.6	201	4 AAG90372	Aag90372 C glutami
31	70	28.6	201	6 ABU25724	Abu25724 Protein e
32	69	28.2	268	6 ADB09362	Adb09362 Alloiococ
33	69	28.2	271	6 ADB09364	Adb09364 Alloiococ
34	68	27.8	203	5 ABB53590	Abb53590 Lactococc
35	68	27.8	205	6 ABU36264	Abu36264 Protein e
36	67	27.3	36	3 AAY77873	Aay77873 B. stearo
37	67	27.3	200	6 ABU43374	Abu43374 Protein e
38	67	27.3	201	6 ABU20674	Abu20674 Protein e
39	67	27.3	205	5 ABU35639	Abu35639 Protein e
40	66	26.9	105	5 ABP03533	Abp03533 Human ORF
41	66	26.9	200	5 ABB49533	Abb49533 Listeria
42	66	26.9	200	6 ABU32565	Abu32565 Protein e
43	66	26.9	202	6 ABU48948	Abu48948 Protein e
44	66	26.9	277	4 AAU45822	Aau45822 Propionib
45	66	26.9	277	6 ABM42341	Abm42341 Propionib

ALIGNMENTS

RESULT 1
AAB48003
ID AAB48003 standard; protein; 251 AA.
XX AC AAB48003;
XX AC
DT 19-MAR-2001 (first entry)
XX AC
DE L. intracellularis hemolysin polypeptide.
XX L
KW Lawsonia intracellularis; immunogenic; hemolysin polypeptide; vaccine;
KW infection; pig; tlyA gene; immune response; antibacterial.
XX L
OS Lawsonia intracellularis.
XX L
FN WO2000069906-A1.
XX L
PD 23-NOV-2000.
XX L
PF 11-MAY-2000, 2000WO-AU000439.
XX L
PR 13-MAY-1999; 99US-0134022P.
XX L

PA (PFIZ) PFIZER PROD INC.
PA (AGRI-) AGRIC VICTORIA SERVICES PTY LTD.
PA (PIGR-) PIG RES & DEV CORP.
XX
PI Panaccio M, Rosey EL, Hasse D, Ankenbauer RG;
XX
XX WPI; 2001-016212/02.
XX N-PSDB; AAC84149.
PT New immunogenic Lawsonia hemolysin peptide, nucleic acid and antibody,
PT useful in vaccines and for the diagnosis of Lawsonia infections,
PT especially in swine.
PS Claim 13; Page 81-82; 86pp; English.
XX

CC The present sequence represents an immunogenic hemolysin polypeptide
CC encoded by a novel gene (tlyA) derived from Lawsonia intracellularis.
CC Recombinant hemolysin polypeptides are used as antigens in vaccines to
CC prevent or treat infection by Lawsonia, in birds and animals, especially
CC pigs, to raise specific antibodies (Ab) and to detect past or present
CC infection. Ab are also useful in diagnosis, to detect L. intracellularis
CC or immunologically cross-reactive species, also for identification of
CC epitopes in hemolysin. Vectors that contain the encoding nucleic acid are
CC also useful in genetic vaccines, and fragments of the tlyA gene are
CC useful as primers or probes for detecting L. intracellularis or related

CC microorganisms, in hybridization or amplification assays

XX Sequence 251 AA;

SQ Query Match 100.0%; Score 245; DB 4; Length 251;
Best Local Similarity 100.0%; Pred. No. 7.2e-26;
Matches 50; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAKHKVRADELVFLQGLAESRQAKELIMAGKVTLTNNSTTIPRLRLEKPG 50
|||||
Db 1 MAKHKVRADELVFLQGLAESRQAKELIMAGKVTLTNNSTTIPRLRLEKPG 50

RESULT 2

AAW55094
ID AAW55094 standard; protein; 270 AA.

XX AC AAW55094;

XX DT 02-OCT-1998 (first entry)

XX Streptococcus pneumoniae SP0041 protein.

XX Streptococcus pneumoniae; antigen; vaccine; infection; diagnosis;
KW detection; pneumonia; otitis media; meningitis.

XX Streptococcus pneumoniae.

XX WO9818930-A2.

XX 07-MAY-1998.

XX 30-OCT-1997; 97WO-US019422.

XX 31-OCT-1996; 96US-0029960P.

XX (HUMA-) HUMAN GENOME SCI INC.

XX Kunsch CA, Choi GH, Johnson LS, Hromockyj A;

XX WPI; 1998-272224/24.

XX N-PSDB; AAV27355.

XX Nucleic acid encoding antigenic peptide(s) from Streptococcus pneumoniae
PT - or their epitope-containing fragments, useful in protective or
PT therapeutic vaccines, and for diagnosis.

XX Claim 11; Page 61; 118pp; English.

XX The present sequence represents a protein from Streptococcus pneumoniae.
CC The nucleic acid sequence encoding the Streptococcus pneumoniae protein
CC can be useful in vaccines for inducing protective antibodies against
CC Streptococcus pneumoniae, for treatment or prevention of infection e.g.
CC pneumonia, otitis media or meningitis. Probes based on the nucleic acid
CC are used to detect Streptococcus infection (by usual hybridisation or
CC amplification methods), also for isolating Streptococcus genes or their
CC allelic variants. The protein can be used similarly to detect specific
CC antibodies in standard immunoassays, especially for diagnosing or
CC monitoring infections. Antibodies which bind the protein are used to
CC detect corresponding antigens, to purify the protein and for passive
CC immunisation (optionally coupled to a toxin). Vaccines are administered,
CC e.g. by injection, orally or through the skin, typically at 0.01-1000
CC (especially 10-300) mu g/ml per dose

XX Sequence 270 AA;

Query Match 38.8%; Score 95; DB 2; Length 270;
Best Local Similarity 50.0%; Pred. No. 0.0001;
Matches 23; Conservative 4; Mismatches 15; Indels 4; Gaps 1;

QY 5 KVRADLVFLQGLAESRQAKELIMAGKVTLTNNSTTIPRLRLEKPG 50
|||||

Db 2 KERVDVLAYKQGLFETREQAKRGWAGLVAVLNGE----RFDKPG 43
|||||

RESULT 3

ABP54588
ID ABP54588 standard; protein; 270 AA.

XX AC ABP54588;

XX DT 04-SEP-2002 (first entry)

XX S. pneumoniae SP041 protein sequence SEQ ID NO:64.

XX Streptococcus pneumoniae; epitope; vaccine; antigenic protein;
KW antibacterial; Streptococcal infection; detection.

XX Streptococcus pneumoniae.

XX US2002061545-A1.

XX 23-MAY-2002.

XX 22-JAN-2001; 2001US-00765272.

XX 30-OCT-1997; 97US-00961083.

XX (CHOI/) CHOI G H.

XX (KUNS/) KUNSCH C A.

XX (BARA/) BARASH S C.

XX (DILL/) DILLON P J.

XX (DOUG/) DOUGHERTY B.

XX (FANN/) FANNON M R.

XX (ROSE/) ROSEN C A.

XX Choi GH, Kunsch CA, Barash SC, Dillon PJ, Dougherty B, Fannon MR;

XX Rosen CA;

XX WPI; 2002-479261/51.

XX N-PSDB; ABQ84823.

XX New Streptococcus pneumoniae antigens, useful for detecting Streptococcus
PT and for preventing or attenuating disease caused by Streptococcus
PT infection.

XX Claim 11; Page 28; 70pp; English.

XX ABQ84792 to ABQ84904 represents nucleic acids which encode the
CC Streptococcus pneumoniae antigens given in ABP54557 to ABP54669. The S.
CC pneumoniae antigens have antibacterial activity and can be used in
CC vaccines. The S. pneumoniae antigens can also be used to prevent or
CC attenuate a Streptococcal infection in an animal. The polynucleotides
CC encoding the S. pneumoniae antigens can be used to detect Streptococcus
CC nucleic acids. ABQ84905 to ABQ85130 represent primers used in the cloning
CC of S. pneumoniae ORFs (open reading frames) which are used in an example
CC from the present invention

XX Sequence 270 AA;

Query Match 38.8%; Score 95; DB 5; Length 270;
Best Local Similarity 50.0%; Pred. No. 0.0001;
Matches 23; Conservative 4; Mismatches 15; Indels 4; Gaps 1;

QY 5 KVRADLVFLQGLAESRQAKELIMAGKVTLTNNSTTIPRLRLEKPG 50
|||||

Db 2 KERVDVLAYKQGLFETREQAKRGWAGLVAVLNGE---RFDKPG 43
|||||

RESULT 4

ADC45145
ID ADC45145 standard; protein; 270 AA.

XX AC ADC45145;

XX DT 18-DEC-2003 (first entry)

```
XX DE S. pneumoniae antigenic protein SP041.
XX KW Antigen; bacterial infection; vaccine; pneumonia; antibacterial.
XX OS Streptococcus pneumoniae.
XX PN US6573082-B1.
XX PD 03-JUN-2003.
XX XX
XX XX 28-MAR-2000; 2000US-00536784.
XX PF 31-OCT-1996; 96US-0029960P.
XX PR 30-OCT-1997; 97US-00961083.
XX XX
XX PA (HUMA-) HUMAN GENOME SCI INC.
XX XX
XX PI Choi GH, Kunsch CA, Barash SC, Dillon PJ, Dougherty B, Fannon MR;
XX PI Rosen CA;
XX DR
XX DR WPI; 2003-764574/72.
XX DR N-PSDB; ADC45144.
XX XX
XX PT Novel polynucleotide encoding Streptococcus pneumoniae polypeptides
XX PT useful for producing vaccines for prevention or attenuation of infection
XX PT by Streptococcus pneumoniae.
XX XX
XX PS Example 1; SEQ ID NO 64; 58pp; English.
XX XX
XX CC The invention relates to an isolated polynucleotide consisting of a
XX CC Streptococcus pneumoniae nucleic acid (appearing as ADC45122 and encoding
XX CC SP028) one of 113 disclosed nucleic acids encoding 113 S. pneumoniae
XX CC antigens. Also included are making a recombinant vector by inserting the
XX CC nucleic acid into a vector, an isolated polynucleotide consisting of at
XX CC least 50 or 100 contiguous nucleotides of the SP028 nucleic acid, and a
XX CC recombinant host cell comprising the SP028 polynucleotide. The nucleic
XX CC acids are useful as DNA vaccine against Streptococcus pneumoniae
XX CC infection (e.g. pneumonia). Nucleic acids derived from the S. pneumoniae
XX CC antigen nucleic acids are useful as probes for use in diagnostic methods
XX CC for detecting S. pneumoniae gene expression. The present sequence
XX CC represents an S. pneumoniae antigenic protein.
XX XX
XX SQ Sequence 270 AA;
Query Match 38.8%; Score 95; DB 7; Length 270;
Best Local Similarity 50.0%; Pred. No. 0.0001;
Matches 23; Conservative 4; Mismatches 15; Indels 4; Gaps 1;
Qy 5 KVRADLVFLQGLAESREQAKRLIMAGKVTLTNNSTTIPRLKPG 50
Db 2 KERVVDVLAYKQGLFETREQAKRGVWAGLVVAVLNGE----RFDKPG 43
RESULT 5
ABP25556
ID ABP25556 standard; protein; 275 AA.
XX AC
XX AC ABP25556;
XX XX
XX DT 02-JUL-2002 (first entry)
XX XX
XX DE Streptococcus polypeptide SEQ ID NO 288.
XX XX
XX KW Streptococcus; GAS; GBS; group B streptococcus; Streptococcus agalactiae;
XX KW group A streptococcus; Streptococcus pyogenes; antibacterial;
XX KW antiinflammatory; infection; vaccine; meningitis; gene therapy.
XX XX
XX OS Streptococcus pyogenes.
XX XX
XX PN WO200234771-A2.
XX XX
XX PD 02-MAY-2002.
```

```
XX XX 29-OCT-2001; 2001WO-GB004789.
XX PF
XX XX 27-OCT-2000; 2000GB-00026333.
XX PR 24-NOV-2000; 2000GB-00028727.
XX PR 07-MAR-2001; 2001GB-00005640.
XX XX
XX PA (CHIR-) CHIRON SPA.
XX PA (GENO-) INST GENOMIC RES.
XX XX
XX PI Telford J, Massignani V, Margarit Y RosI, Grandi G, Fraser C;
XX PI Tettelin H;
XX XX
XX DR WPI; 2002-352536/38.
XX DR N-PSDB; ABN66187.
XX XX
XX PT New Streptococcus protein for the treatment or prevention of infection or
XX PT disease caused by Streptococcus bacteria, such as meningitis, and for
XX PT detecting a compound that binds to the protein.
XX XX
XX PS Claim 1; Page 3183; 4525pp; English.
XX XX
XX CC The invention relates to a protein (ABP25413-ABP30895) from group B
XX CC streptococcus/GBS (Streptococcus agalactiae) or group A streptococcus/GAS
XX CC (Streptococcus pyogenes), comprising one of 5483 sequences (S1), given in
XX CC the specification. The proteins have antibacterial and antiinflammatory
XX CC activity. (I), nucleic acids encoding (I), ABN66044-ABN71526 and
XX CC antibodies that bind (I) are used in the manufacture of medicaments for
XX CC the treatment or prevention of infection or disease caused by
XX CC Streptococcus bacteria, particularly S. agalactiae and S. pyogenes.
XX CC Nucleic acids encoding (I) are used to detect Streptococcus in a
XX CC biological sample. (I) is used to determine whether a compound binds to
XX CC (I). A composition comprising (I) or a nucleic acid encoding (I), may be
XX CC used as a vaccine or diagnostic composition. The disease caused by
XX CC Streptococcus that is prevented or treated may be meningitis. Nucleic
XX CC acid encoding (I) may be used to recombinantly produce (I) and may be
XX CC used in gene therapy. Antibodies to (I) are used for affinity
XX CC chromatography, immunoassays, and distinguishing/identifying
XX CC Streptococcus proteins
XX XX
XX SQ Sequence 275 AA;
Query Match 38.0%; Score 93; DB 5; Length 275;
Best Local Similarity 50.0%; Pred. No. 0.0002;
Matches 23; Conservative 4; Mismatches 15; Indels 4; Gaps 1;
Qy 5 KVRADLVFLQGLAESREQAKRLIMAGKVTLTNNSTTIPRLKPG 50
Db 3 KERVVDVLAYKQGLFETREQAKRGVWAGLVVSVINGQ---RYDKPG 44
RESULT 6
ADC97152
ID ADC97152 standard; protein; 227 AA.
XX AC
XX AC ADC97152;
XX XX
XX DT 01-JAN-2004 (first entry)
XX XX
XX DE E. faecium protein sequence SEQ ID 6779.
XX XX
XX KW Vaccine; urinary tract infection; bacteraemia; endocarditis; wound;
XX KW abdominal-pelvic infection.
XX XX
XX OS Enterococcus faecium.
XX XX
XX PN US6583275-B1.
XX XX
XX PD 24-JUN-2003.
XX XX
XX PF 30-JUN-1998; 98US-00107532.
XX XX
XX PR 02-JUL-1997; 97US-0051571P.
```

```
PR 14-MAY-1998; 98US-0085598P.
XX
XX (GENO-) GENOME THERAPUTICS CORP.
XX
XX Doucette-Stamm LA, Bush D;
XX
XX WPI; 2003-799836/75.
XX N-PSDB; ADC93498.
XX
XX New isolated nucleic acid derived from Enterococcus faecium encoding an
XX Enterococcus faecium polypeptide useful for detection, prevention and
XX treatment of a pathological condition resulting from a bacterial
XX infection.
XX
XX Example 1; SEQ ID NO 6779; 243bp; English.
XX
XX The invention relates to an isolated nucleic acid derived from
XX Enterococcus faecium encoding an Enterococcus faecium polypeptide having
XX one of 10 fully defined sequences given in the (or comprising 40
XX sequential nucleotides chosen from any of the nucleic acids, its
XX complement or sequences hybridising to it). Also included are a
XX recombinant vector comprising the nucleic acid operably linked to
XX transcription regulatory element, a cell comprising the vector and a
XX single-stranded probe comprising the nucleic acid. The nucleic acids are
XX chosen from 3654 disclosed sequences encoding 3654 disclosed proteins.
XX The nucleic acids is useful for diagnosing pathological conditions
XX resulting from E. faecium bacterial infection (e.g. urinary tract
XX infection), bacteraemia, endocarditis, wounds and abdominal-pelvic
XX infection) and for screening drugs such as agonists and antagonists. The
XX nucleic acid is useful for recombinant production of Candida albicans -
XX derived peptides or antisense polypeptides. Pharmaceutical compositions
XX and vaccines containing the nucleic acid are useful for preventing or
XX treating Enterococcus faecium infections. The present sequence represents
XX one if the disclosed E. faecium proteins.
XX
XX Sequence 227 AA;
SQ
Query Match 36.9%; Score 90.5; DB 7; Length 227;
Best Local Similarity 47.8%; Pred. No. 0.00035;
Matches 22; Conservative 4; Mismatches 15; Indels 5; Gaps 1;
QY 5 KVRDELVFLQGLAESREQAQLIMAGKVTLTNNSTTIPRLKPG 50
Db 10 KERVDVLAVKQGLFETREQAKRGSMAGLIYNEKNE-----RFDKPG 50
RESULT 7
ABP25555
ID ABP25555 standard; protein; 275 AA.
XX
XX ABP25555;
XX
XX 02-JUL-2002 (first entry)
XX
XX Streptococcus polypeptide SEQ ID NO 286.
XX
XX Streptococcus; GAS; GBS; group B streptococcus; Streptococcus agalactiae;
XX group A streptococcus; Streptococcus pyogenes; antibacterial;
XX antiinflammatory; infection; vaccine; meningitis; gene therapy.
XX
XX Streptococcus agalactiae.
XX
XX WO200234771-A2.
XX
XX 02-MAY-2002.
XX
XX 29-OCT-2001; 2001WO-GB004789.
XX
XX 27-OCT-2000; 2000GB-00026333.
XX 24-NOV-2000; 2000GB-00028727.
XX 07-MAR-2001; 2001GB-00005640.
XX
XX (CHIR-) CHIRON SPA.
PA
(GENO-) INST GENOMIC RES.
XX
XX Telford J, Masignani V, Margarit Y Rosi, Grandi G, Fraser C;
XX Tettelin H;
XX
XX WPI; 2002-352536/38.
XX N-PSDB; ABN66186.
XX
XX New Streptococcus protein for the treatment or prevention of infection or
XX disease caused by Streptococcus bacteria, such as meningitis, and for
XX detecting a compound that binds to the protein.
XX
XX Claim 1; Page 3183; 4525pp; English.
XX
XX The invention relates to a protein (ABP25413-ABP30895) from group B
XX streptococcus/GBS (Streptococcus agalactiae) or group A streptococcus/GAS
XX (Streptococcus pyogenes), comprising one of 5483 sequences (S1), given in
XX the specification. The proteins have antibacterial and antiinflammatory
XX activity. (I), nucleic acids encoding (I), ABN66044-ABN71526 and
XX antibodies that bind (I) are used in the manufacture of medicaments for
XX the treatment or prevention of infection or disease caused by
XX Streptococcus bacteria, particularly S. agalactiae and S. pyogenes.
XX Nucleic acids encoding (I) are used to detect Streptococcus in a
XX biological sample. (I) is used to determine whether a compound binds to
XX (I). A composition comprising (I) or a nucleic acid encoding (I), may be
XX used as a vaccine or diagnostic composition. The disease caused by
XX Streptococcus that is prevented or treated may be meningitis. Nucleic
XX acid encoding (I) may be used to recombinantly produce (I) and may be
XX used in gene therapy. Antibodies to (I) are used for affinity
XX chromatography, immunoassays, and distinguishing/identifying
XX Streptococcus proteins
XX
XX Sequence 275 AA;
SQ
Query Match 36.7%; Score 90; DB 5; Length 275;
Best Local Similarity 48.0%; Pred. No. 0.00052;
Matches 24; Conservative 6; Mismatches 14; Indels 6; Gaps 2;
QY 1 MAKHKYRADLVFLQGLAESREQAQLIMAGKVTLTNNSTTIPRLKPG 50
Db 1 MAKERV--DVLAYKQGLFDTREQAKRGVMAGVINVINGE---RYDKPG 44
RESULT 8
ABB48304
ID ABB48304 standard; protein; 274 AA.
XX
XX ABB48304;
XX
XX 05-FEB-2002 (first entry)
XX
XX Listeria monocytogenes protein #1008.
XX
XX Antibacterial; gene therapy; vaccine; biosynthesis; biodegradation;
XX vitamin B12; bacterial infection; disease.
XX
XX Listeria monocytogenes.
XX
XX WO200177335-A2.
XX
XX 18-OCT-2001.
XX
XX 11-APR-2001; 2001WO-FR001118.
XX
XX 11-APR-2000; 2000FR-00004629.
XX
XX (INSP ) INST PASTEUR.
XX
XX Buchrieser C, Frangeul L, Couve E, Rusniok C, Fsihi H, Dehoux P;
XX Dussurget O, Chetouani F, Nedjari H, Glaser P, Kunst F, Cossart P;
XX Daniels J, Goebel W, Kreft J, Kuhn M, Ng E, Vazquez-Boland JA;
XX Dominguez-Bernal G, Garrido-Garcia P, Tierrez-Martinez A, Amend A;
XX Chakraborty T, Domann E, Hain T, Berche P, Charbit A, Durant I;
XX
```

PI Perez-Diaz J, Baquero F, Garcia Del Portillo F, Gomez-Lopez N;
PI Maduenio E, De Pablo B, Wehland J, Kaerst U, Entian K, Hauf J;
XX Rose M, Voss H;
XX WPI; 2002-010914/01.
XX Genomic sequence for *Listeria monocytogenes*, useful e.g. for treatment
PT and prevention of *Listeria* and related bacterial infections, and related
PT polypeptides.
XX Claim 6; SEQ ID NO 1009; 192bp; French.
XX The present invention relates to the genome sequence of *Listeria*
CC *monocytogenes* EGD-e (see ABA03041). The genome sequence and fragments of
CC it are useful for selecting probes and primers for detecting genes in *L.*
CC *monocytogenes* and related organisms, and for studying genetic
CC polymorphisms and other genomes. The present sequence is a protein
CC encoded by the genome sequence of the present invention. Proteins
CC expressed from the genome sequence are useful for raising specific
CC antibodies, identification of *L. monocytogenes* and related organisms, and
CC for biosynthesis and biodegradation, especially biosynthesis of Vitamin
CC B12. The genome sequence and proteins encoded by it are also useful for
CC selecting compounds that regulate gene expression and cell replication
CC and modulate *L. monocytogenes*-related diseases. In addition, the genome
CC sequence and proteins encoded by it are useful in pharmaceutical and
CC vaccines compositions for the treatment or prevention of infections by *L.*
CC *monocytogenes* and related organisms. Note: The sequence data for this
CC patent did not form part of the printed specification, but was obtained
CC in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
XX Sequence 274 AA;
Query Match 35.9%; Score 88; DB 5; Length 274;
Best Local Similarity 48.0%; Pred. No. 0.00099;
Matches 24; Conservative 5; Mismatches 15; Indels 6; Gaps 1;
QY 1 MAKHKVADDELIVFTQGLAESREQAKRLIMAGKVTLTNNSTTIPRLKPG 50
Db 1 MTIKKERADILLVEQGLFETREKAKRAIMAGIVYRKSE-----RVDKPG 44
RESULT 9
ABU25523
ID ABU25523 standard; protein; 191 AA.
XX
XX ABU25523;
XX 19-JUN-2003 (first entry)
XX Protein encoded by Prokaryotic essential gene #11050.
XX Antisense; prokaryotic essential gene; cell proliferation; drug design.
XX *Clostridium difficile*.
XX W0200277183-A2.
XX 03-OCT-2002.
XX 21-MAR-2002; 2002WO-US009107.
XX 21-MAR-2001; 2001US-00815242.
PR 06-SEP-2001; 2001US-00948993.
PR 25-OCT-2001; 2001US-0342923P.
PR 08-FEB-2002; 2002US-00072851.
PR 06-MAR-2002; 2002US-0362699P.
XX (ELIT-) ELITRA PHARM INC.
XX Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;
PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;
XX

DR WPI; 2003-0299926/02.
DR N-PSDB; ACA29393.
XX New antisense nucleic acids, useful for identifying proteins or screening
PT for homologous nucleic acids required for cellular proliferation to
PT isolate candidate molecules for rational drug discovery programs.
XX Claim 25; SEQ ID NO 53447; 1766pp; English.
XX The invention relates to an isolated nucleic acid comprising any one of
CC the 6213 antisense sequences given in the specification where expression
CC of the nucleic acid inhibits proliferation of a cell. Also included are:
CC (1) a vector comprising a promoter operably linked to the nucleic acid
CC encoding a polypeptide whose expression is inhibited by the antisense
CC nucleic acid; (2) a host cell containing the vector; (3) an isolated
CC polypeptide or its fragment whose expression is inhibited by the
CC antisense nucleic acid; (4) an antibody capable of specifically binding
CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
CC proliferation or the activity of a gene in an operon required for
CC proliferation; (7) identifying a compound that influences the activity of
CC the gene product or that has an activity against a biological pathway
CC required for proliferation, or that inhibits cellular proliferation; (8)
CC identifying a gene required for cellular proliferation or the biological
CC pathway in which a proliferation-required gene or its gene product lies
CC or a gene on which the test compound that inhibits proliferation of an
CC organism acts; (9) manufacturing an antibiotic; (10) profiling a
CC compound's activity; (11) a culture comprising strains in which the gene
CC product is overexpressed or underexpressed; (12) determining the extent
CC to which each of the strains is present in a culture or collection of
CC strains; or (13) identifying the target of a compound that inhibits the
CC proliferation of an organism. The antisense nucleic acids are useful for
CC identifying proteins or screening for homologous nucleic acids required
CC for cellular proliferation to isolate candidate molecules for rational
CC drug discovery programs, or for screening homologous nucleic acids
CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,
CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of
CC the target prokaryotic essential genes. Note: The sequence data for this
CC patent did not form part of the printed specification, but was obtained
CC in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
XX Sequence 191 AA;
Query Match 35.5%; Score 87; DB 6; Length 191;
Best Local Similarity 40.9%; Pred. No. 0.00087;
Matches 18; Conservative 8; Mismatches 18; Indels 0; Gaps 0;
QY 7 RADELVFLQGLAESREQAKRLIMAGKVTLTNNSTTIPRLKPG 50
Db 83 RLDNVVYRMGLASSRKEARQLVTHGFTLNGNKVDIPSLIVKVG 126
RESULT 10
AAAY77897
ID AAAY77897 standard; protein; 66 AA.
XX
XX AAAY77897;
XX 12-SEP-2003 (revised)
DT 13-JUN-2000 (first entry)
XX
XX B. *stearothermophilus* ribosomal S4 protein fragment (residues 94-159).
DE Ribosomal S4 protein; tRNA anticodon stem-loop; tyrosyl-tRNA synthetase;
KW plant toxin; antibacterial; virucide; antifungal.
XX
XX *Geobacillus stearothermophilus*.
XX W0200011141-A2.
XX 02-MAR-2000.
XX 24-AUG-1999; 99WO-CA000779.
PF

XX 24-AUG-1998; 98US-0097670P.
XX (MOUN) MOUNT SINAI HOSPITAL.
XX Hogue C;
XX WFI; 2000-256376/22.
XX tRNA binding domain, used to determine whether a nucleic acid comprises a
XX tRNA anticodon loop, and in an antibacterial agent, anti-viral agent,
XX immunotoxin, or plant toxin.
XX Claim 6; Fig 2; 57pp; English.
XX The invention provides peptides comprising domains that can bind to tRNA,
XX useful for determining if a DNA contains a tRNA anticodon stem-loop. The
XX substantially pure tRNA binding domain peptide comprises the generic
XX sequence as shown in AAY77898. The peptides are derived from ribosomal S4
XX protein or tyrosyl-tRNA synthetases. The peptides can be used in methods
XX (i) to determine whether a nucleic acid comprises a tRNA anticodon stem-
XX loop; (ii) for determining whether a test compound is an (ant)agonist of
XX a tRNA synthetase-tRNA anticodon stem-loop interaction or a ribosomal S4
XX protein-tRNA anticodon stem-loop interaction; (iii) for interfering with
XX peptide interaction; (iv) to modulate protein synthesis; (v) used in an
XX antibacterial agent, anti-viral agent, immunotoxin, or plant toxin. The
XX peptides can be used to produce antibodies which are immunospecific for
XX the peptides or complexes, and can be used to screen for proteins
XX containing the tRNA anticodon stem-loop recognition motif. The peptides,
XX complexes and antibodies can be used to treat or prevent infections
XX caused by bacteria such as Bacillus, E. coli, Mycobacterium, Hemophilus,
XX Helicobacter, Streptococcus and Staphylococcus, and infections caused by
XX fungi such as S. cerevisiae and Aspergillus. The peptide mimetics are
XX more economical to produce, have greater chemical stability, enhanced
XX pharmaceutical properties such as half-life, absorption, potency, and
XX efficacy, altered specificity, e.g. broad spectrum biological activities,
XX and reduced antigenicity. Sequences AAY77851-897 represent ribosomal S4
XX protein or tyrosyl-tRNA synthetase fragments from various species
XX comprising the domains that can bind tRNA. (Updated on 12-SEP-2003 to
XX standardise OS field)
SQ Sequence 66 AA;

Query Match 31.4%; Score 77; DB 3; Length 66;
Best Local Similarity 38.1%; Pred. No. 0.0059;
Matches 16; Conservative 8; Mismatches 18; Indels 0; Gaps 0;

QY 9 DELVFLQGLAESREQAKRLIMAGKVTLTNNSTTIPLRLKPG 50
DB 2 DNLVYELGLARTERQARQLVTNGHILVDGSRVNIPSYRVKPG 43

RESULT 11
ABP65873
ID ABP65873 standard; protein; 280 AA.
XX ABP65873;
XX 19-NOV-2002 (first entry)
XX Bifidobacterium longum NCC2705 ORF amino acid sequence SEQ ID NO:617.
XX Bifidobacterium longum NCC2705; Bifidobacterium; bacterial;
XX anti-diarrheic; antibacterial; inhibitor of Salmonella; detection;
XX identification; lactic acid bacterium; diarrhoea; pathogenic bacteria;
XX rotavirus; food composition; pharmaceutical composition.
XX Bifidobacterium longum.
XX RP127152-A1.
XX 31-JUL-2002.
XX

PF 30-JAN-2001; 2001EP-00102050.
XX 30-JAN-2001; 2001EP-00102050.
XX (NEST) SOC PROD NESTLE SA.
XX WFI; 2002-668397/72.
XX Novel polynucleotide comprising Bifidobacterium genome sequence useful as
XX a probe or primer for detecting and/or identifying Bifidobacterium longum
XX in a biological sample.
XX Claim 3; SEQ ID NO 617; 80pp; English.
XX

XX The present invention describes a polynucleotide (I) comprising a
XX sequence of a Bifidobacterium genome selected from the nucleotide
XX sequences given in AB081842 and AB081843, or a sequence exhibiting at
XX least 90% identity or which hybridises with the sequences given in
XX AB081842 and AB081843. Also described is a polynucleotide (II) encoding a
XX fusion protein, comprising a sequence selected from 1097 sequences given
XX in ABP65258 to ABP66354 ligated in frame to a polynucleotide encoding a
XX heterologous polypeptide. (I) has anti-diarrheic and antibacterial
XX activities, and can be used as an inhibitor of Salmonella. (I) (which is
XX a probe) is useful for the detection and/or identification of
XX Bifidobacterium longum in a biological sample. A carrier containing the
XX lactic acid bacterium Bifidobacterium longum NCC2705 (NCIM 1-2618) can be
XX used for preventing and/or treating diarrhoea brought about by pathogenic
XX bacteria and/or rotavirus. The carrier is a food composition selected
XX from milk, yogurt, curd, cheese, fermented milks, milk based fermented
XX products, ice-creams, fermented cereal based products, milk based
XX powders, infant formula, pet food or a pharmaceutical composition
XX selected from tablets, liquid bacterial suspensions, dried oral
XX supplement, wet oral supplement, dry tube feeding or wet tube feeding.
XX (I) is useful in DNA arrays or chips to carry out analysis of the
XX expression of the Bifidobacterium gene. ABQ81844 to ABQ81850 represent
XX Bifidobacterium related nucleotide sequences given in the Sequence
XX Listing from the present invention but not mentioned further within the
XX specification. N.B. the sequence data for this patent is not represented
XX in the printed specification but is based on sequence information
XX supplied by the European Patent Office
SQ Sequence 280 AA;

Query Match 31.0%; Score 76; DB 5; Length 280;
Best Local Similarity 40.9%; Pred. No. 0.05;
Matches 18; Conservative 7; Mismatches 19; Indels 0; Gaps 0;

QY 7 RADELVFLQGLAESREQAKRLIMAGKVTLTNNSTTIPLRLKPG 50
DB 37 RLDMMLVAGSLVESRAKQRLIKAGHVRVDGETITRPSFMVKAG 80

RESULT 12
ADB11952
ID ADB11952 standard; protein; 139 AA.
XX ADB11952;
XX 20-NOV-2003 (first entry)
XX Alloicoccus otitis antigenic protein SEQ ID NO:5108.
XX Alloicoccus otitis; antigenic protein; immunogenic; immunisation;
XX gene therapy; Gram-positive bacterium; infection.
XX Alloicoccus otitis.
XX WO2003048304-A2.
XX 12-JUN-2003.
XX 25-NOV-2002; 2002WO-US036123.
XX


```

PR 29-NOV-2001; 2001US-0333777P.
PR 18-NOV-2002; 2002US-0426742P.
XX (AMHP ) WYETH HOLDINGS CORP.
XX
XX Fletcher LD, McMichael JC, Russell DP, Zagursky RJ;
XX
XX WPI; 2003-505284/47.
XX N-PSDB; ADB11955.
XX
XX New Alloicoccus otitidis polynucleotides and polypeptides, useful for
XX treating and diagnosing diseases, drug screening assays and monitoring of
XX effects during drug clinical trials.
XX
XX Claim 33; SEQ ID NO 5108; 1019pp; English.
XX
XX The present invention describes an isolated polynucleotide (I) of
XX Alloicoccus otitidis genomic DNA, which encodes an antigenic protein.
XX Alloicoccus otitidis is a Gram-positive bacterium. Also described: (1)
XX an isolated polypeptide that is encoded by the polynucleotide (I); (2) an
XX expression vector comprising the novel isolated polynucleotide (I), its
XX complement, degenerate variant or fragment; (3) a genetically engineered
XX host cell, transfected, transformed or infected with the vector of (2);
XX (4) an antibody specific for the polypeptide of (1); (5) an immunogenic
XX composition comprising the polypeptide, its complement, biological
XX equivalent or fragment, or the polynucleotide that is comprised in the
XX expression vector; (6) a pharmaceutical composition comprising the
XX polypeptide of (1) and a carrier; (7) a protein chip comprising an array
XX of the polypeptides of (1), their biological equivalent or fragment; (8)
XX immunising against Alloicoccus otitidis by administering to a host the
XX immunogenic composition; (9) detecting and/or identifying Alloicoccus
XX otitidis in the biological sample; (10) a kit comprising a container
XX containing the novel polynucleotide, its degenerate variant or fragment,
XX or the antibody of (4); and (11) producing a polypeptide by culturing the
XX genetically engineered host cell under conditions suitable to produce the
XX polypeptide from the culture. (I) can be used in gene therapy. The
XX polynucleotides, polypeptides, antibodies and compositions of the present
XX invention can be used for treating and diagnosing diseases, drug
XX screening assays and monitoring of effects during drug clinical trials.
XX The polynucleotides are useful for expressing and detecting Alloicoccus
XX otitidis. The present sequence represents an Alloicoccus otitidis
XX antigen protein from the present invention.
XX
XX SQ Sequence 139 AA;
XX
XX Query Match 30.6%; Score 75; DB 6; Length 139;
XX Best Local Similarity 37.8%; Pred. No. 0.029;
XX Matches 14; Conservative 9; Mismatches 14; Indels 0; Gaps 0;
XX
XX QY 7 RADELVFLOGLASREQAOKPLIMAGKVTLTNNSTTIP 43
XX | | | | | | | | | | | | | | | | | |
XX 32 RLDNIVYRLGLASTRRQARQLVVGHTITVDGKRVDP 68
XX
XX RESULT 13
XX ID ADB11950 standard; protein; 201 AA.
XX AC ADB11950;
XX
XX DT 20-NOV-2003 (first entry)
XX
XX DE Alloicoccus otitis antigenic protein SEQ ID NO:5110.
XX
XX KW Alloicoccus otitidis; antigenic protein; immunogenic; immunisation;
XX gene therapy; Gram-positive bacterium; infection.
XX
XX OS Alloicoccus otitis.
XX
XX WO2003048304-A2.
XX
XX PD 12-JUN-2003.
XX
XX
XX PF 25-NOV-2002; 2002WO-US036123.
XX
XX PR 29-NOV-2001; 2001US-0333777P.
XX PR 18-NOV-2002; 2002US-0426742P.
XX
XX XX (AMHP ) WYETH HOLDINGS CORP.
XX
XX Fletcher LD, McMichael JC, Russell DP, Zagursky RJ;
XX
XX WPI; 2003-505284/47.
XX N-PSDB; ADB11953.
XX
XX New Alloicoccus otitidis polynucleotides and polypeptides, useful for
XX treating and diagnosing diseases, drug screening assays and monitoring of
XX effects during drug clinical trials.
XX
XX Claim 33; SEQ ID NO 5110; 1019pp; English.
XX
XX The present invention describes an isolated polynucleotide (I) of
XX Alloicoccus otitidis genomic DNA, which encodes an antigenic protein.
XX Alloicoccus otitidis is a Gram-positive bacterium. Also described: (1)
XX an isolated polypeptide that is encoded by the polynucleotide (I); (2) an
XX expression vector comprising the novel isolated polynucleotide (I), its
XX complement, degenerate variant or fragment; (3) a genetically engineered
XX host cell, transfected, transformed or infected with the vector of (2);
XX (4) an antibody specific for the polypeptide of (1); (5) an immunogenic
XX composition comprising the polypeptide, its complement, biological
XX equivalent or fragment, or the polynucleotide that is comprised in the
XX expression vector; (6) a pharmaceutical composition comprising the
XX polypeptide of (1) and a carrier; (7) a protein chip comprising an array
XX of the polypeptides of (1), their biological equivalent or fragment; (8)
XX immunising against Alloicoccus otitidis by administering to a host the
XX immunogenic composition; (9) detecting and/or identifying Alloicoccus
XX otitidis in the biological sample; (10) a kit comprising a container
XX containing the novel polynucleotide, its degenerate variant or fragment,
XX or the antibody of (4); and (11) producing a polypeptide by culturing the
XX genetically engineered host cell under conditions suitable to produce the
XX polypeptide from the culture. (I) can be used in gene therapy. The
XX polynucleotides, polypeptides, antibodies and compositions of the present
XX invention can be used for treating and diagnosing diseases, drug
XX screening assays and monitoring of effects during drug clinical trials.
XX The polynucleotides are useful for expressing and detecting Alloicoccus
XX otitidis. The present sequence represents an Alloicoccus otitidis
XX antigen protein from the present invention.
XX
XX SQ Sequence 201 AA;
XX
XX Query Match 30.6%; Score 75; DB 6; Length 201;
XX Best Local Similarity 37.8%; Pred. No. 0.045;
XX Matches 14; Conservative 9; Mismatches 14; Indels 0; Gaps 0;
XX
XX QY 7 RADELVFLOGLASREQAOKPLIMAGKVTLTNNSTTIP 43
XX | | | | | | | | | | | | | | | | | |
XX 94 RLDNIVYRLGLASTRRQARQLVVGHTITVDGKRVDP 130
XX
XX RESULT 14
XX ID ABG98423 standard; protein; 209 AA.
XX AC ABG98423;
XX
XX DT 16-JAN-2003 (first entry)
XX
XX DE Thermus thermophilus 30S ribosomal protein S4.
XX
XX KW Ribosome; three dimensional structure; Thermus thermophilus;
XX x-ray crystallography; protein coordinate data.
XX
XX OS Thermus thermophilus.
XX
XX WO200246392-A2.
XX
XX

```

(SMIK) SMITHKLINE BEECHAM CORP.
(SMIK) SMITHKLINE BEECHAM PLC.

Black MT, Hodgson JE, Knowles DJC, Reichard RW, Nicholas RO;
Burnham MKR, Pratt JM, Rosenberg M, Ward JM, Lonetto MA;

WPI; 1998-252940/23.
N-PSDB; AAV53464.

New nucleic acid sequences from Staphylococcus aureus WCHU29 - useful in vaccines and for treatment of bacterial infections of e.g. respiratory tract and central nervous system.

Claim 11; Page 320; 390pp; English.

This sequence represents a Staphylococcus aureus protein, that based on homology with a Bacillus subtilis protein, is a 30S ribosomal protein S4 (Bs4), and is encoded by a DNA sequence of the invention. The DNA sequences were isolated from Staphylococcus aureus WCHU29 (NCIME 40771). Host cells containing the DNA sequences are used to produce polypeptides or fragments. The proteins are used in the treatment of disease, for inducing an immune response by administering them, to produce antibody and/or T-cell immune response. Antagonists of the proteins are used for the inhibition of bacterial polypeptides. Conditions which may be treated include bacterial infections, especially respiratory, cardiac, gastrointestinal, central nervous, eye, kidney, urinary tract, skin, bones and joints. The proteins can also be used to identify antimicrobial compounds which are broad spectrum antibiotics, especially useful in the treatment of H. pylori infection

Sequence 137 AA;

Query Match 30.2%; Score 74; DB 2; Length 137;
Best Local Similarity 36.4%; Pred. No. 0.039;
Matches 16; Conservative 8; Mismatches 20; Indels 0; Gaps 0

QY 7 RADELVFLQGLAESREQAKRLIMAGKVTLTNNSTTIPLRKPG 50
| | : | : | : | : | : | : | : | : | : | : | : | : | :
Db 14 RLDVVVSLGLARTRRQARQLVNHGHLVDGKRVDIPSYVKPG 57

RESULT 16
AAV70150
ID AAV70150 standard; protein; 195 AA.
AC AAV70150;
XX XX
DT 06-JUN-2000 (first entry)
DE Staphylococcus aureus protein (HGS034).
XX Vaccine; antibacterial; prevention; attenuation; detection;
KW Staphylococcal infection; neonatal conjunctivitis; skin infection;
KW toxic shock syndrome; osteomyelitis.
XX Staphylococcus aureus.
OS
XX Key Location/Qualifiers
FH Domain 33..35
FT /label= Antigenic epitope
XX XX
PN WO200012678-A2.
XX PD
XX 09-MAR-2000.
XX 31-AUG-1999; 99WO-US019726.
XX PF
XX 01-SEP-1998; 98US-0098964P.
XX PR
XX (HUMA-) HUMAN GENOME SCI INC.
PA
XX Bailey CC, Choi GH;
PI

13-JUN-2002.
XX
XX PA
XX PF
XX XX
XX PR
XX PR
XX PR
XX XX
XX PA
XX XX
XX Noller HF, Cate JHD, Yusupov MM, Yusupova GZ, Baucom AE;
XX Lancaster L, Dallas A, Lieberman K;
XX WPI; 2002-713237/77.
XX
XX Novel three dimensional structure of Thermus thermophilus 70S ribosome resolved using x-ray crystallography upto 5.5 Angstroms resolution, useful for screening and designing compounds that alter ribosome function.
XX
XX Claim 8; Page 512-513; 525pp; English.
XX
XX This invention relates to a three dimensional structure of Thermus thermophilus 70S ribosome resolved using x-ray crystallography up to 5.5 Angstrom resolution. The invention also discloses a method for identifying a compound that binds to a bacterial 70S ribosome or its portion. The method comprises designing a compound based upon a three-dimensional structure of the bacterial 70S ribosome, where the structure co-ordinates are given in the specification, contacting the compound with the bacterial 70S ribosome or its portion and determining whether the compound binds to the bacterial 70S ribosome or its portion. The 70S ribosome 3D structure of the invention can be used to screen and design compounds that bind to sites on the 70S ribosome and that alter the ribosome function. The present sequence represents a ribosomal protein sequence used in the method of the invention

Sequence 209 AA;

Query Match 30.6%; Score 75; DB 5; Length 209;
Best Local Similarity 34.1%; Pred. No. 0.048;
Matches 15; Conservative 9; Mismatches 20; Indels 0; Gaps 0;

QY 7 RADELVFLQGLAESREQAKRLIMAGKVTLTNNSTTIPLRKPG 50
| | : | : | : | : | : | : | : | : | : | : | : | : | :
Db 100 RLDNVVYLGFVSRQRQLRVHGHITVNGRRVDLPYRVVRPG 143

RESULT 15
AAW77671
ID AAW77671 standard; protein; 137 AA.
XX AC
XX AAW77671;
XX AC
DT 30-OCT-1998 (first entry)
XX DE
XX 30S ribosomal protien S4.
XX Staphylococcus aureus protein; immune response induction; eye infection;
XX antibody production; T-cell immune response; gastrointestinal infection;
KW respiratory infection; inhibitor; bacterial infection; cardiac infection;
KW central nervous system; kidney infection; urinary tract infection;
KW antimicrobial compound identification; broad spectrum antibiotic;
therapy.
XX
XX Staphylococcus aureus.
OS
XX EF841394-A2.
XX PN
XX 13-MAY-1998.
XX PD
XX 24-SEP-1997; 97BP-00307485.
XX PF
XX 24-SEP-1996; 96US-0027032P.
XX RR

CC aureus polypeptides. The invention is useful for assaying staphylococcal
CC infection in an animal by assaying expression of genes encoding
CC staphylococcal polypeptides. The invention incorporated in a biosensor is
CC useful for detecting, diagnosing, and monitoring *S. aureus* or other
CC staphylococcus sp. and their infections, to detect particular
CC polynucleotides, and to monitor the genetic changes in response to drug
CC therapy in the clinic and drug development. The present sequence is
CC Staphylococcus aureus HGS034
XX
XX Sequence 195 AA;
SQ
Query Match 30.2%; Score 74; DB 5; Length 195;
Best Local Similarity 36.4%; Pred. No. 0.06; Indels 0; Gaps 0;
Matches 16; Conservative 8; Mismatches 20;
QY 7 RADELVFLQGLAESRQAKRLIMAGKVTLTNNSTTIPRLRLEKPG 50
DB 93 RLDVVYSLGLARTRRQARQLVNHGHILVDGKRVDPISYVKPG 136
RESULT 18
AAU37185
ID AAU37185 standard; protein; 200 AA.
XX
AC AAU37185;
XX
DT 14-FEB-2002 (first entry)
XX
DE Staphylococcus aureus cellular proliferation protein #1355.
XX
KW Antisense; prokaryotic cellular proliferation protein; antibiotic;
KW antibacterial; drug design.
XX
OS Staphylococcus aureus.
XX
PN WO200170955-A2.
XX
PD 27-SEP-2001.
XX
PF 21-MAR-2001; 2001WO-US009180.
XX
PR 21-MAR-2000; 2000US-0191078P.
PR 23-MAY-2000; 2000US-0206848P.
PR 26-MAY-2000; 2000US-0207727P.
PR 23-OCT-2000; 2000US-0242578P.
PR 27-NOV-2000; 2000US-0253625P.
PR 22-DEC-2000; 2000US-0257911P.
PR 16-FEB-2001; 2001US-0269308P.
XX
PA (ELIT-) ELITRA PHARM INC.
XX
PI Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;
PI Yamamoto RT, Xu HH;
XX
DR WPI; 2001-611495/70.
DR N-PSDB; AAS55044.
XX
PT New polynucleotides for the identification and development of
PT antibiotics, comprise sequences of antisense nucleic acids.
XX
XX Example 3; SEQ ID NO 12778; 511pp; English.
XX
XX The invention relates to antisense inhibitors of genes essential to
XX prokaryotic cellular proliferation, their use in identifying the genes,
XX their use in the discovery of novel antibiotics, the essential genes
XX themselves and the encoded proteins. The prokaryotes used are *Escherichia*
XX coli, *Staphylococcus aureus*, *Salmonella typhi*, *Klebsiella pneumoniae*,
XX *Pseudomonas aeruginosa* and *Enterococcus faecalis*. The invention is also
XX useful for the identification of potential new targets for antibiotic
XX development. The antisense nucleic acids can also be used to identify
XX proteins used in proliferation, to express these proteins, and to obtain
XX antibodies capable of binding to the expressed proteins. The proteins can
XX be used to screen compounds in rational drug discovery programmes. The

XX
DR WPI; 2000-237864/20.
DR N-PSDB; AAZ51233.
XX
PT Staphylococcus aureus polypeptide useful for preventing or attenuating a
PT staphylococcal infection comprises one of 32 sequences of 100-1277 amino
PT acids or their fragments.
XX
XX Claim 9; Page 19; 144pp; English.
XX
XX The present sequence is a protein (HGS034) from Staphylococcus aureus
XX genomic DNA library. The present sequence is useful in preparation of
XX vaccines for prevention or attenuation of Staphylococcal infections
XX (especially *S. aureus* infections) which may cause conditions such as
XX neonatal conjunctivitis, osteomyelitis, skin infections and toxic shock
XX syndrome. The present sequence is also useful for detecting
XX Staphylococcal infections in biological samples
XX
XX Sequence 195 AA;
SQ
Query Match 30.2%; Score 74; DB 3; Length 195;
Best Local Similarity 36.4%; Pred. No. 0.06;
Matches 16; Conservative 8; Mismatches 20; Indels 0; Gaps 0;
QY 7 RADELVFLQGLAESRQAKRLIMAGKVTLTNNSTTIPRLRLEKPG 50
DB 93 RLDVVYSLGLARTRRQARQLVNHGHILVDGKRVDPISYVKPG 136
RESULT 17
AAE25699
ID AAE25699 standard; protein; 195 AA.
XX
AC AAE25699;
XX
DT 04-NOV-2002 (first entry)
XX
DE Staphylococcus aureus HGS034.
XX
KW Recombinant vector; staphylococcal infection; genetic change; enzyme;
KW drug therapy; HGS034.
XX
OS Staphylococcus aureus.
XX
FH Key Location/Qualifiers
FH 33..35
FT Region /note="Antigenic epitope"
FT
XX
PN US6403337-B1.
XX
XX 11-JUN-2002.
PD
XX 24-FEB-2000; 2000US-00512255.
PF
XX 03-JAN-1997; 97US-00781986.
PR 20-OCT-1997; 97US-00956171.
PR 31-AUG-1999; 99WO-US019726.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Bailey C, Choi GH;
PI
XX WPI; 2002-536035/57.
DR N-PSDB; AAD42223.
XX
XX Novel Staphylococcus aureus genes useful for generating recombinant
PT vectors and for detecting, diagnosing, and monitoring Staphylococcus
PT aureus or other Staphylococcus species infections.
XX
XX Disclosure; Col 37-38; 123pp; English.
PS
XX The invention relates to novel Staphylococcus aureus genes. The invention
XX is useful for producing a recombinant vector, for use in producing *S.*

CC antisense nucleic acid sequence is also useful to screen for homologous
CC nucleic acids which are required for cell proliferation in a wide variety
CC of organisms. The present sequence represents an essential prokaryotic
CC cellular proliferation protein. Note: The sequence data for this patent
CC did not form part of the printed specification, but was obtained in
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
XX Sequence 200 AA;

Query Match 30.2%; Score 74; DB 4; Length 200;
Best Local Similarity 36.4%; Pred. No. 0.062;
Matches 16; Conservative 8; Mismatches 20; Indels 0; Gaps 0;

QY 7 RADELVFLQGLAESRQAKRLIMAGKVTLTNNSTTIPRLKPG 50
| | | | | | | | | | | | | | | | | | | | | |
Db 93 RLDAVYSLGLARTRQARQLVNHGHILVDGKRVDPISYVRPG 136

RESULT 19

AAU33908
ID AAU33908 standard; protein; 200 AA.

XX
AC AAU33908;

XX
DT 14-FEB-2002 (first entry)

XX
DE Staphylococcus aureus cellular proliferation protein #184.

XX
KW Antisense; prokaryotic cellular proliferation protein; antibiotic;
KW antibacterial; drug design.

XX
OS Staphylococcus aureus.

XX
FN WO200170955-A2.

XX
PD 27-SEP-2001.

XX
PF 21-MAR-2001; 2001WO-US009180.

XX
PR 21-MAR-2000; 2000US-0191078P.

XX
PR 23-MAY-2000; 2000US-0206848P.

XX
PR 26-MAY-2000; 2000US-0207727P.

XX
PR 23-OCT-2000; 2000US-0242578P.

XX
PR 27-NOV-2000; 2000US-0253625P.

XX
PR 22-DEC-2000; 2000US-0257931P.

XX
PR 16-FEB-2001; 2001US-0269308P.

XX
FA (ELIT-) ELITRA PHARM INC.

XX
PI Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;
PI Yamamoto RT, Xu HH;

XX
XX WPI; 2001-611495/70.

XX
DR N-PSDB; AAS51767.

XX
PT New polynucleotides for the identification and development of
PT antibiotics, comprise sequences of antisense nucleic acids.

XX
PS Example 3; SEQ ID NO 5404; 51lpp; English.

XX
CC The invention relates to antisense inhibitors of genes essential to
CC prokaryotic cellular proliferation, their use in identifying the genes,
CC their use in the discovery of novel antibiotics, the essential genes
CC themselves and the encoded proteins. The prokaryotes used are Escherichia
CC coli, Staphylococcus aureus, Salmonella typhi, Klebsiella pneumoniae,
CC Pseudomonas aeruginosa and Enterococcus faecalis. The invention is also
CC useful for the identification of potential new targets for antibiotic
CC development. The antisense nucleic acids can also be used to identify
CC proteins used in proliferation, to express these proteins, and to obtain
CC antibodies capable of binding to the expressed proteins. The proteins can
CC be used to screen compounds in rational drug discovery programmes. The
CC antisense nucleic acid sequence is also useful to screen for homologous

CC nucleic acids which are required for cell proliferation in a wide variety
CC of organisms. The present sequence represents an essential prokaryotic
CC cellular proliferation protein. Note: The sequence data for this patent
CC did not form part of the printed specification, but was obtained in
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
XX Sequence 200 AA;

Query Match 30.2%; Score 74; DB 4; Length 200;
Best Local Similarity 36.4%; Pred. No. 0.062;
Matches 16; Conservative 8; Mismatches 20; Indels 0; Gaps 0;

QY 7 RADELVFLQGLAESRQAKRLIMAGKVTLTNNSTTIPRLKPG 50
| | | | | | | | | | | | | | | | | | | | | |
Db 93 RLDAVYSLGLARTRQARQLVNHGHILVDGKRVDPISYVRPG 136

RESULT 20

AAU36985
ID AAU36985 standard; protein; 200 AA.

XX
AC AAU36985;

XX
DT 14-FEB-2002 (first entry)

XX
DE Staphylococcus aureus cellular proliferation protein #1155.

XX
KW Antisense; prokaryotic cellular proliferation protein; antibiotic;
KW antibacterial; drug design.

XX
OS Staphylococcus aureus.

XX
PN WO200170955-A2.

XX
PD 27-SEP-2001.

XX
PF 21-MAR-2001; 2001WO-US009180.

XX
PR 21-MAR-2000; 2000US-0191078P.

XX
PR 23-MAY-2000; 2000US-0206848P.

XX
PR 26-MAY-2000; 2000US-0207727P.

XX
PR 23-OCT-2000; 2000US-0242578P.

XX
PR 27-NOV-2000; 2000US-0253625P.

XX
PR 22-DEC-2000; 2000US-0257931P.

XX
PR 16-FEB-2001; 2001US-0269308P.

XX
FA (ELIT-) ELITRA PHARM INC.

XX
PI Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;
PI Yamamoto RT, Xu HH;

XX
XX WPI; 2001-611495/70.

XX
DR N-PSDB; AAS54844.

XX
PT New polynucleotides for the identification and development of
PT antibiotics, comprise sequences of antisense nucleic acids.

XX
PS Example 3; SEQ ID NO 12578; 51lpp; English.

XX
CC The invention relates to antisense inhibitors of genes essential to
CC prokaryotic cellular proliferation, their use in identifying the genes,
CC their use in the discovery of novel antibiotics, the essential genes
CC themselves and the encoded proteins. The prokaryotes used are Escherichia
CC coli, Staphylococcus aureus, Salmonella typhi, Klebsiella pneumoniae,
CC Pseudomonas aeruginosa and Enterococcus faecalis. The invention is also
CC useful for the identification of potential new targets for antibiotic
CC development. The antisense nucleic acids can also be used to identify
CC proteins used in proliferation, to express these proteins, and to obtain
CC antibodies capable of binding to the expressed proteins. The proteins can
CC be used to screen compounds in rational drug discovery programmes. The
CC antisense nucleic acid sequence is also useful to screen for homologous
CC nucleic acids which are required for cell proliferation in a wide variety

Protein encoded by Prokaryotic essential gene #1637.

Antisense; prokaryotic essential gene; cell proliferation; drug design.

Staphylococcus aureus.

WO200277183-A2.

03-OCT-2002.

21-MAR-2002; 2002WO-US009107.

21-MAR-2001; 2001US-00815242.

06-SEP-2001; 2001US-00948993.

25-OCT-2001; 2001US-0342923P.

08-FEB-2002; 2002US-00072851.

06-MAR-2002; 2002US-0362693P.

(ELIT-) ELITRA PHARM INC.

Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;
Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;

WPI: 2003-029926/02.

N-PSDB; ACA19980.

New antisense nucleic acids, useful for identifying proteins or screening for homologous nucleic acids required for cellular proliferation to isolate candidate molecules for rational drug discovery programs.

Claim 25; SEQ ID NO 44034; 1766pp; English.

The invention relates to an isolated nucleic acid comprising any one of the 6213 antisense sequences given in the specification where expression of the nucleic acid inhibits proliferation of a cell. Also included are: (1) a vector comprising a promoter operably linked to the nucleic acid encoding a polypeptide whose expression is inhibited by the antisense nucleic acid; (2) a host cell containing the vector; (3) an isolated polypeptide or its fragment whose expression is inhibited by the antisense nucleic acid; (4) an antibody capable of specifically binding the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular proliferation or the activity of a gene in an operon required for proliferation; (7) identifying a compound that influences the activity of the gene product or that has an activity against a biological pathway required for proliferation, or that inhibits cellular proliferation; (8) identifying a gene required for cellular proliferation or the biological pathway in which a proliferation-required gene or its gene product lies or a gene on which the test compound that inhibits proliferation of an organism acts; (9) manufacturing an antibiotic; (10) profiling a compound's activity; (11) a culture comprising strains in which the gene product is overexpressed or underexpressed; (12) determining the extent to which each of the strains is present in a culture or collection of strains; or (13) identifying the target of a compound that inhibits the proliferation of an organism. The antisense nucleic acids are useful for identifying proteins or screening for homologous nucleic acids required for cellular proliferation to isolate candidate molecules for rational drug discovery programs, or for screening homologous nucleic acids required for proliferation in cells other than *S. aureus*, *S. typhimurium*, *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of the target prokaryotic essential genes. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at [ftp.wipo.int/pub/published_pct_sequences](http://wipo.int/pub/published_pct_sequences)

Sequence 200 AA;

Query Match 30.2%; Score 74; DB 6; Length 200;
Best Local Similarity 36.4%; Pred. No. 0.062;
Matches 16; Conservative 8; Mismatches 20; Indels 0; Gaps 0;

7 RADELVFLQGLAFSRQAKRLIMAGKVTLLTNSTTILRLKPG 50
93 RLDAVYSLGLARTRQQLVNHGLVQKRVDPISYVKPG 136

[illegible]

Db 94 RLDNVFRLGLATRRQARQFVNHGHLVDGKRVDPSPFRVQPG 137

RESULT 35

RESUL 33
ABU36264
ID ABU36264 standard; protein: 205 AA.

XX
AC ABU36264:

XX
DT 19-JUN-2003 (first entry)

XX
DE Protein encoded by Prokaryotic essential gene #21791.

XX Antisense; prokaryotic essential gene; cell proliferation; drug design.

XX *Mycoplasma pneumoniae*.

XX PN WO200277183-A2.

XX
PD 03-OCT-2002.

XX
PF
21-MAR-2002; 2002WO-US009107.

XX
PR 21-MAR-2001: 2001US-00815242.

PR 06-SEP-2001; 2001US-00948993.
PR 25-OCT-2001; 2001US-0342923P.

PR 08-FEB-2002; 2002US-00072851.
PR 06-MAR-2002; 2002US-0362699P.

PA (ELITT-) ELITRA PHARM INC
XX
XX

XX Wang L., Zamudio C. Malone C
PT (SEE) , SEE PAGE FURTHER ON.

PI Wall D, Trawick JD, Car
XX

DR WPI: 2003-029926/02.
DR N-PSDB: ACA40134.

XX PT New antisense nucleic acids, useful for identifying proteins or screening

PT for homologous nucle
PT isolate candidate mo

XX
PS Claim 25; SEQ ID

XX The invention relates to an isolated nucleic acid molecule encoding a protein, wherein the protein is a protein that is involved in the regulation of the cell cycle.

CC the 6213 antisense sequences given in the specification where expression
CC of the nucleic acid inhibits proliferation of a cell. Also included are:

CC (1) a vector comprising a promoter operably linked to the nucleic acid
CC encoding a polypeptide whose expression is inhibited by the antisense

CC nucleic acid; (2) a host cell containing the
CC polypeptide or its fragment whose expression

CC antisense nucleic acid; (4) an antibody capable of specifically binding
CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular

CC proliferation or the activity of a gene in an operon required for
CC proliferation; (7) identifying a compound that influences the activity of
CC proliferation;

CC the gene product or that has an activity against a biological pathway
CC required for proliferation, or that inhibits cellular proliferation; (8)

CC identifying a gene required for cellular proliferation or the biological
CC pathway in which a proliferation-required gene or its gene product lies

CC or a gene on which the test compound that inhibits proliferation of an
CC organism acts; (9) manufacturing an antibiotic; (10) profiling a

CC compound's activity; (11) a culture comprising strains in which the gene
CC product is overexpressed or underexpressed; (12) determining the extent

CC to which each of the strains is present in a culture or collection of
CC strains; or (13) identifying the target of a compound that inhibits the

CC proliferation of an organism. The antisense nucleic acids are useful for
CC identifying proteins or screening for homologous nucleic acids required

CC for cellular proliferation to isolate candidate molecules for rational
CC drug discovery programs, or for screening homologous nucleic acids

CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,
CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of

the target prokaryotic essential genes. Note: The sequence data for this patent did not form part of the printed specification, but was obtained

PT New antisense nucleic acids, useful for identifying proteins or screening
PT for homologous nucleic acids required for cellular proliferation to
PT isolate candidate molecules for rational drug discovery programs.
XX
PS Claim 25; SEQ ID NO 63563; 1766pp; English.

XX (CURA-) CURAGEN CORP.
PA Shimkets RA, Leach MD;
XX WPI; 2002-106308/14.
XX N-PSDB; ABN19285.
XX Novel human polypeptides and polynucleotides useful for diagnosing,
PT preventing and treating cardiovascular disease, neurodegenerative,
PT hyperproliferative disorders and autoimmune disorders.
XX Disclosure; SEQ ID NO 7048; 1037pp; English.
XX
CC The present invention describes substantially purified human proteins
CC (referred to as open reading frame, ORFX, where X is 1-11491 (see Table 1
CC in the specification). ABN15762 to ABN27252 encode the human ORFX
CC proteins given in ABP00010 to ABP11500. ORFX proteins are useful for
CC treating or preventing a pathology associated with an ORFX-associated
CC disorder in humans, and in the manufacture of a medicament for treating a
CC syndrome associated with ORFX-associated disorder. ORFX polynucleotide
CC sequences can be used in gene therapy. ORFX sequences can be used in the
CC treatment of cancer, hyperproliferative disorders, cirrhosis of liver,
CC psoriasis, benign tumours, keloid, degenerative disorders, haemorrhage,
CC osteoarthritis, neurodegenerative disorders, disorders related to organ
CC transplantation, cardiovascular diseases, diabetes mellitus, systemic
CC lupus erythematosus, hypertension, hypothyroidism, cholesterol ester
CC storage disease, various immune deficiencies and disorders, infectious
CC diseases, autoimmune disorders such as multiple sclerosis, rheumatoid
CC arthritis, autoimmune thyroiditis, myasthenia gravis, graft-versus-host
CC disease and autoimmune inflammatory eye disease. ORFX proteins are also
CC useful for treating burns, incisions, ulcers, for treating osteoporosis,
CC bone degenerative disorders, or periodontal disease, and for gut
CC protection or regeneration and treatment of lung or liver fibrosis,
CC reperfusion injury in various tissues and conditions resulting from
CC systemic cytokine damage. N.B. The sequence data for this patent did not
CC form part of the printed specification, but was obtained in electronic
CC format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 105 AA;

Query Match 26.9%; Score 66; DB 5; Length 105;
Best Local Similarity 42.1%; Pred. No. 0.37;
Matches 16; Conservative 5; Mismatches 17; Indels 0; Gaps 0;

Qy 6 VRDELVFLQGLAESREQAKRLIMAGKVTLTNNSTIP 43
| | | | | | | | | | | | | | | | | | | | | |
Db 19 VRLDQALVEAGLARSRTLATRLVREGRTVNRVVVTKP 56

Search completed: July 7, 2004, 16:54:45
Job time : 18.9435 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 7, 2004, 16:52:58 ; Search time 27.5183 seconds
(without alignments)
470.891 Million cell updates/sec

Title: US-10-009-919A-1

Perfect score: 1271

Sequence: 1 MAKHKVRADELVFLQGLAES.....PSVIKPKGNQYLIYLKKR 251

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA:*

- 1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep:*
- 2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:*
- 3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep:*
- 4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep:*
- 5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep:*
- 6: /cgn2_6/ptodata/2/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	446.5	35.1	270	3	US-08-961-083-64
2	446.5	35.1	270	4	US-09-536-784-64
3	433.5	34.1	276	4	US-09-634-238-259
4	377.5	29.7	240	1	US-07-965-668A-3
5	377.5	29.7	240	2	US-08-950-433-3
6	377.5	29.7	240	3	US-09-186-287-3
7	365	28.7	227	4	US-09-107-532A-6779
8	310.5	24.4	219	4	US-09-134-000C-6791
9	89.5	7.0	304	4	US-09-328-352-7082
10	87.5	6.9	397	4	US-09-252-991A-2235
11	87.5	6.9	415	4	US-09-207-388-12
12	87.5	6.9	562	4	US-09-207-388-14
13	87.5	6.9	642	4	US-09-207-388-13
14	87.5	6.9	642	4	US-09-207-388-15
15	87.5	6.9	662	4	US-09-207-388-16
16	86	6.8	536	4	US-09-107-532A-6930
17	84.5	6.6	160	1	US-08-307-279A-12
18	84.5	6.6	160	4	US-09-525-310-12
19	84	6.6	281	4	US-09-134-000C-5454
20	83	6.5	321	4	US-09-328-352-4883
21	83	6.5	464	3	US-08-957-063-3
22	83	6.5	464	4	US-09-487-685-3
23	83	6.5	464	4	US-08-802-805D-3
24	83	6.5	464	4	US-09-187-906-13
25	83	6.5	664	3	US-08-957-063-16
26	83	6.5	664	4	US-09-487-685-16
27	83	6.5	664	4	US-08-802-805D-16

28	83	6.5	2125	4	US-09-919-172-29	Sequence 29, Appl
29	82.5	6.5	340	4	US-09-328-352-5334	Sequence 5334, Ap
30	81	6.4	464	3	US-08-957-063-6	Sequence 6, Appli
31	81	6.4	464	4	US-09-487-685-6	Sequence 6, Appli
32	81	6.4	464	4	US-08-802-805D-6	Sequence 6, Appli
33	81	6.4	464	3	US-08-957-063-18	Sequence 18, Appl
34	81	6.4	664	4	US-09-487-685-18	Sequence 18, Appl
35	81	6.4	664	4	US-08-802-805D-18	Sequence 18, Appl
36	80.5	6.3	406	4	US-09-134-001C-3570	Sequence 3570, Ap
37	80	6.3	276	4	US-09-328-352-6056	Sequence 6056, Ap
38	80	6.3	287	4	US-09-540-236-2477	Sequence 2477, Ap
39	80	6.3	607	2	US-08-472-534-5	Sequence 5, Appli
40	80	6.3	638	4	US-09-489-039A-8996	Sequence 8996, Ap
41	79.5	6.3	320	1	US-08-465-980-2	Sequence 2, Appli
42	79.5	6.3	320	2	US-09-053-303-2	Sequence 2, Appli
43	79.5	6.3	320	4	US-09-339-115-2	Sequence 2, Appli
44	79.5	6.3	320	5	PCT-US95-07093-2	Sequence 2, Appli
45	79.5	6.3	447	3	US-09-109-204-3	Sequence 3, Appli

ALIGNMENTS

RESULT 1
US-08-961-083-64
; Sequence 64, Application US/08961083
; Patent No. 6159469
; GENERAL INFORMATION:
; APPLICANT: Choi et. al.
; TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines
; NUMBER OF SEQUENCES: 452
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850

COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/961,083
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Brookes, A. Anders
; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: PB340P2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 64:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 270 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-961-083-64

Query Match 35.1%; Score 446.5; DB 3; Length 270;
Best Local Similarity 42.7%; Pred. No. 1.6e-46;
Matches 106; Conservative 41; Mismatches 94; Indels 7; Gaps 4;
5 KVRADLVFQGLASREQAKLIMAGKYLITNNSTTILRLKPKGHKYPLEISICSLGV 64
2 KERVVDVLYAKQGLFETREQAKGVAGLVAVVINGE----RFDKPKGKIPDDTELKKG 57


```
Query Match          24.4%; Score 310.5; DB 4; Length 219;
Best Local Similarity 44.5%; Pred. No. 7.3e-30;
Matches 81; Conservative 26; Mismatches 72; Indels 3; Gaps 3;

QY 69 SRGAYKLLTALDFPKIDVKSCICLDAGASTGGFTDCLLOHGASKVYAYDVKGQGLHEKLY 128
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 11 SWGGLKLEKALNFAINVQGTMLDGSSTGFTDVALQNGARLSYALDVGYNQLAWKIR 70
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 129 TNEQVINIEGVNLTASKDLIPREV-DILITDVFSISLTLLPSCIRWLKASGIIIIALIK 187
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 71 QDERVVMERTNFRYSKPEDFTGVPDIATIDVFSILRLILPLHDLKKGGSWALIK 130
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 188 POPELYPDKI-KKGVVKETSQYEAIVEKLIHFQCOSELGLIFIGVVPVSKPGKGNQBYLI 246
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 131 PQFAGKEFGKGIIVRDPETHQMVVEIITRFAMNN-GYDVKNLDFSPITGGEGNIEFLA 189
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 247 YL 248
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 190 HL 191
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 9
US-09-328-352-7082
; Sequence 7082, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breston et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
; TITLE OF INVENTION: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC99-03PA
; CURRENT APPLICATION NUMBER: US/09/328,352
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 7082
; LENGTH: 304
; TYPE: PRT
; ORGANISM: Acinetobacter baumannii
US-09-328-352-7082

Query Match          7.0%; Score 89.5; DB 4; Length 304;
Best Local Similarity 24.8%; Pred. No. 0.025;
Matches 32; Conservative 19; Mismatches 53; Indels 25; Gaps 3;

QY 83 KIDVKSICLDAGASTGGFTDCLLOHGASKVYAYDVKGQGLHEKLYTNEQVINIEGVNLR 142
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 158 KTDVKDKIVIDYGCSGILGVAALLGAKKVATDIDP-----QAVIATKQNAELGVLDLR 213
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 143 -----TASKDLIPREVDILITDVFSISLTLLPSCIRWLKASGIIIIALIKPOPELYPD 195
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 214 LYVGLPEFDDQEFKQADVLVANILAGPLMALAPEFAKLKSDG-----D 259
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 196 KIKGVVKE 204
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 260 PALAGVIEE 268
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 10
US-09-252-991A-22235
; Sequence 22235, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 22235
```

```
; LENGTH: 397
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-22235

Query Match          6.9%; Score 87.5; DB 4; Length 397;
Best Local Similarity 23.8%; Pred. No. 0.069;
Matches 34; Conservative 25; Mismatches 41; Indels 43; Gaps 7;

QY 44 LRLEKPGHKYPLSISCSLIGVERFVSR--GAYKLLTAL-----DFFKIDVKSCI 90
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 148 VRAEKE-YGVPAETIIIVSIIIGVETIFFGRTGTSYRMDALSTLIGDYPFPRAFFRKELREFL 206
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 91 CL-----DAGASTGGFTDCL-----LOHGASKVYAYDV-KGQQLH----- 124
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 207 LLAREQOVDPLSLTGSYAGAMGLPQFMPSPSFRAYAVDFDGDGHINWSDPTDAIGSVASY 266
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 125 ---EKLYTNEQVINIEGVNLR 144
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 267 FKQHGVTGEPVVSVAEINDESA 289
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 11
US-09-207-388-12
; Sequence 12, Application US/09207388
; Patent No. 6497880
; GENERAL INFORMATION:
; APPLICANT: Wisniewski, Jan
; TITLE OF INVENTION: HEAT SHOCK GENES AND PROTEINS FROM
; TITLE OF INVENTION: NEISSERIA MENINGITIDIS, CANDIDA GLABRATA AND ASPERGILLUS
; FILE REFERENCE: 870109.411
; CURRENT APPLICATION NUMBER: US/09/207,388
; CURRENT FILING DATE: 1998-12-08
; NUMBER OF SEQ ID NOS: 102
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 12
; LENGTH: 415
; TYPE: PRT
; ORGANISM: Neisseria meningitidis
US-09-207-388-12

Query Match          6.9%; Score 87.5; DB 4; Length 415;
Best Local Similarity 23.7%; Pred. No. 0.074;
Matches 62; Conservative 35; Mismatches 86; Indels 79; Gaps 13;

QY 10 ELVFLOGLAESREQAKRLIMAGKVTLTN-----NSTTIPRLKPKGHKVPLESICSLICV 64
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 194 DVMALQRLKEAAEKAKIELSSGQQQTEINLPYITMDATGPKHLAMKITRAKPESL-----V 248
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 65 ERFVSRGAYKLLTALDFFKIDVKSCICLDAGASTGGFTDCLLOHGASKVYAYDVKGQQLH 124
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 249 EDLITR-----STEPCKIALK-----DAGLSTGDIIDVILVGGQSRMPKV----- 288
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 125 EKLYTNEQVINIEGVNLTASKDLIPREV-----DILITDVFSISLT 167
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 289 -----QEAVKAPFGKEPR---KDVNPDDEAVAGAAIQGEVLSSGGRSDVLLLDVTLPLSLGI 340
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 168 ILPSCIRWLKASGIIIIALIKQFELYPKIKKGVVKETSLQYEAIVEKLIHFQCOSE----- 222
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 341 -----ETMGGVMTKLIQKNTTI-PTKASQ-VFSTAEDNQSAV--TIHVLQGERERAS 388
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 223 ----LGLIFIGVVPVSKPGK 240
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 389 ANKSLGQFNLG---DIAPAPRG 407
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 12
US-09-207-388-14
; Sequence 14, Application US/09207388
; Patent No. 6497880
; GENERAL INFORMATION:
; APPLICANT: Wisniewski, Jan
```

; TITLE OF INVENTION: HEAT SHOCK GENES AND PROTEINS FROM
; TITLE OF INVENTION: NEISSERIA MENINGITIDIS, CANDIDA GLABRATA AND ASPERGILLUS
; TITLE OF INVENTION: FUMIGATUS
; FILE REFERENCE: 870109.411
; CURRENT APPLICATION NUMBER: US/09/207,388
; CURRENT FILING DATE: 1998-12-08
; NUMBER OF SEQ ID NOS: 102
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 14
; LENGTH: 562
; TYPE: PRT
; ORGANISM: Neisseria meningitidis
US-09-207-388-14

Query Match 6.9%; Score 87.5; DB 4; Length 562;
Best Local Similarity 23.7%; Pred. No. 0.15;
Matches 62; Conservative 35; Mismatches 86; Indels 79; Gaps 13;
QY 10 ELVFLOGLAESRQAKRLIMAGKVTLTN-----NSTTIPRLKPGHKYPLESICSLIGV 64
Db 176 DVNALQRLKEAAEKAKIELSSGQQTETNLPYITMDATGPKHLAMKITRAKFESL-----V 230
QY 65 ERFVSRGAYKLLTALDFFKIDVKSCICLDAGASTGGFTDCLLOHGASKVVAIDVGKQLH 124
Db 231 EDLITR-----SIEPCKIALK-----DAGLSTGDDIDVILVGGQSRMPKV----- 270
QY 125 EKLYTNEQVINIEGVNLRASKDLIPEV-----DILTIDVSFISLTL 167
Db 271 -----QEAVKAFPGKEPR---KDVNPDDEAVAGAAIQGEVLSGRSDVLLDVTPLSLGI 322
QY 168 ILPSCIRWLKASGIIIALIKPQELYPDKIKKGVVKETSLOYEAVEKIIHFCQSE----- 222
Db 323 -----ETWGGVMTKLIQKNTTI--PTKASQ--VFSTADNQSAV--TIHVLQGERERAS 370
QY 223 ----LGLIFIGVVPVSIKPGK 240
Db 371 ANKSLGQFNLG---DIAPAPRG 389

RESULT 13
US-09-207-388-13
; Sequence 13, Application US/09207388
; Patent No. 6497880
; GENERAL INFORMATION:
; APPLICANT: Wisniewski, Jan
; TITLE OF INVENTION: HEAT SHOCK GENES AND PROTEINS FROM
; TITLE OF INVENTION: NEISSERIA MENINGITIDIS, CANDIDA GLABRATA AND ASPERGILLUS
; FILE REFERENCE: 870109.411
; CURRENT APPLICATION NUMBER: US/09/207,388
; CURRENT FILING DATE: 1998-12-08
; NUMBER OF SEQ ID NOS: 102
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 13
; LENGTH: 642
; TYPE: PRT
; ORGANISM: Neisseria meningitidis
US-09-207-388-13

Query Match 6.9%; Score 87.5; DB 4; Length 642;
Best Local Similarity 23.7%; Pred. No. 0.15;
Matches 62; Conservative 35; Mismatches 86; Indels 79; Gaps 13;
QY 10 ELVFLOGLAESRQAKRLIMAGKVTLTN-----NSTTIPRLKPGHKYPLESICSLIGV 64
Db 256 DVNALQRLKEAAEKAKIELSSGQQTETNLPYITMDATGPKHLAMKITRAKFESL-----V 310
QY 65 ERFVSRGAYKLLTALDFFKIDVKSCICLDAGASTGGFTDCLLOHGASKVVAIDVGKQLH 124
Db 311 EDLITR-----SIEPCKIALK-----DAGLSTGDDIDVILVGGQSRMPKV----- 350
QY 125 EKLYTNEQVINIEGVNLRASKDLIPEV-----DILTIDVSFISLTL 167

Db 351 -----QEAVKAFPGKEPR---KDVNPDDEAVAGAAIQGEVLSGRSDVLLDVTPLSLGI 402
QY 168 ILPSCIRWLKASGIIIALIKPQELYPDKIKKGVVKETSLOYEAVEKIIHFCQSE----- 222
Db 403 -----ETWGGVMTKLIQKNTTI--PTKASQ--VFSTADNQSAV--TIHVLQGERERAS 450
QY 223 ----LGLIFIGVVPVSIKPGK 240
Db 451 ANKSLGQFNLG---DIAPAPRG 469

RESULT 14
US-09-207-388-15
; Sequence 15, Application US/09207388
; Patent No. 6497880
; GENERAL INFORMATION:
; APPLICANT: Wisniewski, Jan
; TITLE OF INVENTION: HEAT SHOCK GENES AND PROTEINS FROM
; TITLE OF INVENTION: NEISSERIA MENINGITIDIS, CANDIDA GLABRATA AND ASPERGILLUS
; FILE REFERENCE: 870109.411
; CURRENT APPLICATION NUMBER: US/09/207,388
; CURRENT FILING DATE: 1998-12-08
; NUMBER OF SEQ ID NOS: 102
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 15
; LENGTH: 642
; TYPE: PRT
; ORGANISM: Neisseria meningitidis
US-09-207-388-15

Query Match 6.9%; Score 87.5; DB 4; Length 642;
Best Local Similarity 23.7%; Pred. No. 0.15;
Matches 62; Conservative 35; Mismatches 86; Indels 79; Gaps 13;
QY 10 ELVFLOGLAESRQAKRLIMAGKVTLTN-----NSTTIPRLKPGHKYPLESICSLIGV 64
Db 256 DVNALQRLKEAAEKAKIELSSGQQTETNLPYITMDATGPKHLAMKITRAKFESL-----V 310
QY 65 ERFVSRGAYKLLTALDFFKIDVKSCICLDAGASTGGFTDCLLOHGASKVVAIDVGKQLH 124
Db 311 EDLITR-----SIEPCKIALK-----DAGLSTGDDIDVILVGGQSRMPKV----- 350
QY 125 EKLYTNEQVINIEGVNLRASKDLIPEV-----DILTIDVSFISLTL 167
Db 351 -----QEAVKAFPGKEPR---KDVNPDDEAVAGAAIQGEVLSGRSDVLLDVTPLSLGI 402
QY 168 ILPSCIRWLKASGIIIALIKPQELYPDKIKKGVVKETSLOYEAVEKIIHFCQSE----- 222
Db 403 -----ETWGGVMTKLIQKNTTI--PTKASQ--VFSTADNQSAV--TIHVLQGERERAS 450
QY 223 ----LGLIFIGVVPVSIKPGK 240
Db 451 ANKSLGQFNLG---DIAPAPRG 469

RESULT 15
US-09-207-388-16
; Sequence 16, Application US/09207388
; Patent No. 6497880
; GENERAL INFORMATION:
; APPLICANT: Wisniewski, Jan
; TITLE OF INVENTION: HEAT SHOCK GENES AND PROTEINS FROM
; TITLE OF INVENTION: NEISSERIA MENINGITIDIS, CANDIDA GLABRATA AND ASPERGILLUS
; FILE REFERENCE: 870109.411
; CURRENT APPLICATION NUMBER: US/09/207,388
; CURRENT FILING DATE: 1998-12-08
; NUMBER OF SEQ ID NOS: 102
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 16
; LENGTH: 662
; TYPE: PRT

FEATURE: NAME/KEY: misc feature
LOCATION: (B) LOCATION 1...536
SEQUENCE DESCRIPTION: SEQ ID NO: 6930:
US-09-107-532A-6930

Query Match 6.8%; Score 86; DB 4; Length 536;
Best Local Similarity 21.9%; Pred. No. 0.17;
Matches 54; Conservative 40; Mismatches 93; Indels 60; Gaps 10;

QY 14 LQGLAESREOAKLIMAGKVTLLNNSTTIPRLKPGHKYPLESICSLIGVRFVSRGAY 73
Db 242 LQRLKDAEAKKOLSG-----VSSQTSLPITAGESPLH-----LEMTLTKAF 288
QY 74 KLLTA--LDPFKIDVKSCICLDAGASTGGTDCLLQHGASKVYVAIDVKGQQLHEKLYTNE 131
Db 289 DEITADIVERTKVPVRQAL-KDAGLSQSEIDVILVGGSTRIPAV----- 332
QY 132 QVINIEGVNLRTA---SKDLIPEV-----DILTIDVSFISLTLLIPSCIRWLKAS 179
Db 333 ---VEAVRKETGKEPNKSVNPDEVVANGAAIQGGVITGDKVLLDVTPLSLGIETMG 388
QY 180 GIITIALIKPOFELYDPDKIKGVVKETSLQYEAVEKLIHFCQSE-----LGLIFIGV 230
Db 389 GVFTKLDIRNTTIPTSKQ--VFSTAADNQPAVD--IHVLOGERPMAADNKTLCGRFQLT 444
QY 231 VPSVIKG 237
Db 445 IPAAPRG 451

RESULT 17
US-08-307-279A-12
US-08-307-279A-12 Application US/08307279A
Sequence 12, Application US/08307279A
Patent No. 5736347
GENERAL INFORMATION:
APPLICANT: Anderson, Burt E.
TITLE OF INVENTION: Nucleic Acids of Rochalimaea Henselae
TITLE OF INVENTION: and Methods and Compositions for Diagnosing Rochalimaea
TITLE OF INVENTION: Henselae and Rochalimaea Quintana Infection
NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESS:
ADDRESSEE: NEEDLE & ROSENBERG, P.C.
STREET: 127 Peachtree Street, N.E., Suite 1200
CITY: Atlanta
STATE: Georgia
COUNTRY: USA
ZIP: 30303
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/307,279A
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Spratt, Gwendolyn D.
REGISTRATION NUMBER: 36,016
REFERENCE/DOCKET NUMBER: 1414.624
TELECOMMUNICATION INFORMATION:
TELEPHONE: (404) 688-0770
TELEFAX: (404) 688-9880
INFORMATION FOR SEQ ID NO: 12:
SEQUENCE CHARACTERISTICS:
LENGTH: 160 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-307-279A-12

ORGANISM: Neisseria meningitidis
US-09-207-388-16

Query Match 6.9%; Score 87.5; DB 4; Length 662;
Best Local Similarity 23.7%; Pred. No. 0.16;
Matches 62; Conservative 35; Mismatches 86; Indels 79; Gaps 13;

QY 10 ELVFLQGLAESROAKLIMAGKVTLLNNSTTIPRLKPGHKYPLESICSLIGV 64
Db 276 DVNALQRLKEAARAKIELSSGQQTINLPYITMDATGPKHLAMKITRAFESL-----V 330
QY 65 ERFVSRGAYKLLTALDPFKIDVKSCICLDAGASTGGTDCLLQHGASKVYVAIDVKGQQLH 124
Db 331 EDLITR-----SIEPCKALK-----DAGLSIGDIDVILVGGQSRMPKV----- 370
QY 125 EKLYTNEQVINIEGVNLRKASDLIPEV-----DILTIDVSFISLTLL 167
Db 371 -----QEAUKAFFGKEPR--KQVNPDEAVAVAGAAIQEVLGSGRSDVLLLDVTPLSLGI 422
QY 168 ILPSCIRWLKASGIIILIKPOFELYDPDKIKGVVKETSLQYEAVEKLIHFCQSE----- 222
Db 423 -----ETMGGYMTKLIQKNTTI-PTKASQ--VFSTAEDNQSAV--TIHVLOGERERAS 470
QY 223 ---LGLIFIGWPSVIKGPKG 240
Db 471 ANKSLGQFNLG---DIAPAPRG 489

RESULT 16
US-09-107-532A-6930
US-09-107-532A, Application US/09107532A
Sequence 6930, Application US/09107532A
Patent No. 6583275
GENERAL INFORMATION:
APPLICANT: Lynn A Doucette-Stamm and David Bush
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS
NUMBER OF SEQUENCES: 7310
CORRESPONDENCE ADDRESS:
ADDRESSEE: GENOME THERAPEUTICS CORPORATION
STREET: 100 Beaver Street
CITY: Waltham
STATE: Massachusetts
COUNTRY: USA
ZIP: 02354
COMPUTER READABLE FORM:
MEDIUM TYPE: CD-ROM ISO9660
COMPUTER: PC
OPERATING SYSTEM: <Unknown>
SOFTWARE: ASCII
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/107,532A
FILING DATE: 30-Jun-1998
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/085,598
FILING DATE: 14 May 1998
APPLICATION NUMBER: 60/051571
FILING DATE: July 2, 1997
ATTORNEY/AGENT INFORMATION:
NAME: Ariniello, Pamela Deneke
REGISTRATION NUMBER: 40,489
REFERENCE/DOCKET NUMBER: GTC-012
TELECOMMUNICATION INFORMATION:
TELEPHONE: (781)893-5007
TELEFAX: (781)893-8277
INFORMATION FOR SEQ ID NO: 6930:
SEQUENCE CHARACTERISTICS:
LENGTH: 536 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: YES
ORIGINAL SOURCE:
ORGANISM: Enterococcus faecium

```

Query Match          6.6%; Score 84.5; DB 1; Length 160;
Best Local Similarity 26.5%; Pred. No. 0.038;
Matches 41; Conservative 24; Mismatches 61; Indels 29; Gaps 8;

QY 73 YKLTALDFEKIDVKSICLDAGASTGGTDCLLQHGASKVVAIDVKGQLHFKLYTNE- 131
Db 16 YSLVTLULSF-----CIS-HAKAQATATLTDEYYKKALENTOKLDVAKSQTAESIYESAT 68

QY 132 -----QVINIGVNLRTASKOLIPREVILITDVSFISLITLILPSCIRWLKASGIIIAL 185
Db 69 QTANKIKDINNQLANLKADTKTK-PEQLAQALQELTLLQALQADT-----LKIQS--LAM 121

QY 186 I-----KPOFELYPDKIKKGVVKETSLSQYEAVEKI 215
Db 122 IQAKDTTKKEELREEQTK---KHEDLQQLKEKL 153

RESULT 18
US-09-525-310-12
; Sequence 12, Application US/09525310
; Patent No. 6406887
; GENERAL INFORMATION:
; APPLICANT: Anderson, Burt E.
; Regnery, Russell L.
; TITLE OF INVENTION: Nucleic Acids of Rochalimaea Henselae
; and Methods and Compositions for Diagnosing Rochalimaea
; Henselae and Rochalimaea Quintana Infection
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NEEDLE & ROSENBERG, P.C.
; STREET: 127 Peachtree Street, N.E., Suite 1200
; CITY: Atlanta
; STATE: Georgia
; COUNTRY: USA
; ZIP: 30303
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/525,310
; FILING DATE: 14-Mar-2000
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/307,279
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Spratt, Gwendolyn D.
; REGISTRATION NUMBER: 36,016
; REFERENCE/DOCKET NUMBER: 1414.624
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (404) 688-0770
; TELEFAX: (404) 688-9880
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 160 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 12:
US-09-525-310-12

```

```

Query Match          6.6%; Score 84.5; DB 4; Length 160;
Best Local Similarity 26.5%; Pred.No. 0.038;
Matches 41; Conservative 24; Mismatches 61; Indels 29; Gaps 8;

QY 73 YKLLTALDFPKIDVKSCICLDAGASTGCFDCLLQHGASKVYADVGCGOLHKEYLNE- 131
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 16 YSLVLLSLF-----CIS-HAKAQTATLDEYKKALENTQKLDVAKSQTAISYSAT 68
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 132 -----QVINEGVLNLTASKDLPIPEVDILTIDVFSISLTLPSCIRMLKASGIIIL 185
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

Db      69  QTANKIDINNLKADTKTK-PEQIQALQIETLLQALQADT-----LKIQS--LAM 121

QY      186  I-----KPQFELYDPDKIRKGVVKETSLOYEAVEKI 215
           : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      122  IQANDTKTEELREBOTK---KHEDLOKQKLEK 153

RESULT 19
US-09-134-000C-5454
; Sequence 5454, Application US/09134000C
; Patent No. 6617156
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; TITLE OF INVENTION: ENTEROCOCCUS FAECALIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 032796-032
; CURRENT APPLICATION NUMBER: US/09/134,000C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/055,778
; PRIOR FILING DATE: 1997-08-15
; NUMBER OF SEQ ID NOS: 6812
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 5454
; LENGTH: 281
; TYPE: PRT
; ORGANISM: Enterococcus faecalis
US-09-134-000C-5454

```

Query Match.	6.6%;	Score 84;	DB 4;	Length 281;
Best Local Similarity	22.5%;	Pred. No. 0.11;		
Matches	54;	Conservative	42;	Mismatches 100; Indels 44; Gaps 10;
QY	16	GLAESREQAKRLMAGKVTLTNNSTTIPLRLEKPGHK-YPLESTCSLLIGVERFVSRGAYK	74	
Db	64	GIAAACQLGTPIWISGNVSIN-----ETDQAITYTPMI-GMWGVIEDVSQ----	109	
QY	75	LTLTALDPFKIDVKSCICLDAGASTGGFTDCLOHGASKVYAIDVGK--GOLH-----EKL	127	
Db	110	-INTQAFKQVDLILYLIGETHADFNG-----SEIQIKIQLRGTEGQLRSFDLKEEK	158	
QY	128	YTNEQVINTEGYNLRLTASKDLIPREVDILTIDVSP---ISLTLLTLCSCPWLKA---	SGI 181	
Db	159	ANQELVLKALQVLGVAASADCAEGGVAVALAESAFANELQVLPDKLEYLFAETSQRF	218	
QY	182	ITALIKPQELVPDKTKK-----GVVKETSQVEAVEKIIFHCOSSELGLFIIGVVPVSVIK	236	
Db	219	IUSVSPHQGEAETTLGRKAQHIGKVTEGLVTHALDDVNCTSKAKALWEIDAIPCMLK	278	

```

RESULT 20
US-09-328-352-4883
; Sequence 4883, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELAT
; FILE REFERENCE: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS
; CURRENT APPLICATION NUMBER: US/09/328,352
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 6252
; SEQ ID NO 4883
; LENGTH: 321
; TYPE: PRT
; ORGANISM: Acinetobacter baumannii
US-09-328-352-4883

```

```
Query Match          6.5%; Score 83; DB 4; Length 321;
Best Local Similarity 28.4%; Pred. No. 0.18;
Matches 38; Conservative 22; Mismatches 50; Indels 24; Gaps 8;

QY      104 CLLCGCAKVV-----AIDVGKGGLHKLVTNEGVNIEGVNLT---ASKDLI-PREVDI 155
```


Db 61 CFLDHAGVPTFPAGNGLDVG---HPHIGLQFTTWMIEGTMMHTDSLGLSKQIRPKQVNL 117
QY 156 LTIDVFSISLTILPSCIRWLKASIIIAL-----LKPOELYPD--KIKKGVVKETSL 207
Db 118 MTAGHG-ISHTVAPDFTETQHAQAQLTALPDHKRNDMPKEHYHFDLPVVEKQGLEFTVL 176
QY 208 QYEAVER---IIH 217
Db 177 VGEIETTSPVVH 190
RESULT 21
US-08-957-063-3
; Sequence 3, Application US/08957063
; Patent No. 6025157
; GENERAL INFORMATION:
; APPLICANT: Robert D. Klein, Arnon Rosenthal, Mary A. Hynes
; TITLE OF INVENTION: Neurturin Receptor
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 1 DNA Way
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WinPatIn (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/957,063
; FILING DATE: 24-Oct-1997
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/871
; FILING DATE: 9-Jun-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 913
; FILING DATE: 18-Feb-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Torchia, PhD., Timothy E.
; REGISTRATION NUMBER: 36,700
; REFERENCE/DOCKET NUMBER: P1086P2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650/225-8674
; TELEFAX: 650/952-9881
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 464 amino acids
; TYPE: Amino Acid
; TOPOLOGY: Linear
US-08-957-063-3
Query Match 6.5%; Score 83; DB 3; Length 464;
Best Local Similarity 22.3%; Pred. No. 0.32; Indels 92; Gaps 13;
Matches 60; Conservative 23; Mismatches 94; Indels 92; Gaps 13;
QY 6 VRADLVFLQGLAESRQAKRLIMAGK---VLTNNSTTIPRL--EXP-----GHK 52
Db 41 VRANELCAESNCSSRYRTLQCLAGDRNTMLANKECAALEVQLQESPLYDCRCRGMK 100
QY 53 YPLESI-----CSLIGVERFVSRGAYKLTA--LDFFK-----IDVKSCI 90
Db 101 KELQCLQIYWSIHLGLTEGEFFYEASPEYPTVTSRLSDIFRLASIFSGTGADPVSASKNH 160
QY 91 CLDAGASTGGTDCCLQHQGASKVYVADVGKQLHEKLYTNEQVINIEGVNLTASK---- 146
Db 161 CLDAKA-----CNLNDCKLRSSYIS-----ICNREISPTERCNRKCHKALRQ 206
QY 147 --DLIPEEVDILTIDVFSISL-----TLILPSCIRWLKASIIIALIKPOFELYP 194

Db 207 FDRVPSE---YTYRMLFCSQDQACAEARRRQTILPSC-----SYE 244
QY 195 DKIKKGVVKETSLQYEAVERKIIHFQCOSEL 223
Db 245 DK-----EKNCLDLRGVCRTHLCRSRL 268
RESULT 22
US-09-487-685-3
; Sequence 3, Application US/09487685
; Patent No. 6342348
; GENERAL INFORMATION:
; APPLICANT: Robert D. Klein, Arnon Rosenthal, Mary A. Hynes
; TITLE OF INVENTION: Neurturin Receptor
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 1 DNA Way
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WinPatIn (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/487,685
; FILING DATE: 19-Jan-2000
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/957,063
; FILING DATE: <Unknown>
; APPLICATION NUMBER: 913
; FILING DATE: 18-Feb-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Torchia, PhD., Timothy E.
; REGISTRATION NUMBER: 36,700
; REFERENCE/DOCKET NUMBER: P1086P2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650/225-8674
; TELEFAX: 650/952-9881
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 464 amino acids
; TYPE: Amino Acid
; TOPOLOGY: Linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 3:
US-09-487-685-3
Query Match 6.5%; Score 83; DB 4; Length 464;
Best Local Similarity 22.3%; Pred. No. 0.32; Indels 92; Gaps 13;
Matches 60; Conservative 23; Mismatches 94; Indels 92; Gaps 13;
QY 6 VRADLVFLQGLAESRQAKRLIMAGK---VLTNNSTTIPRL--EXP-----GHK 52
Db 41 VRANELCAESNCSSRYRTLQCLAGDRNTMLANKECAALEVQLQESPLYDCRCRGMK 100
QY 53 YPLESI-----CSLIGVERFVSRGAYKLTA--LDFFK-----IDVKSCI 90
Db 101 KELQCLQIYWSIHLGLTEGEFFYEASPEYPTVTSRLSDIFRLASIFSGTGADPVSASKNH 160
QY 91 CLDAGASTGGTDCCLQHQGASKVYVADVGKQLHEKLYTNEQVINIEGVNLTASK---- 146
Db 161 CLDAKA-----CNLNDCKLRSSYIS-----ICNREISPTERCNRKCHKALRQ 206
QY 147 --DLIPEEVDILTIDVFSISL-----TLILPSCIRWLKASIIIALIKPOFELYP 194
Db 207 FDRVPSE---YTYRMLFCSQDQACAEARRRQTILPSC-----SYE 244
QY 195 DKIKKGVVKETSLQYEAVERKIIHFQCOSEL 223

Db 245 DK-----EKNCLDLRGVCRTHLCSRL 268

RESULT 23

US-08-802-805D-3
; Sequence 3, Application US/08802805D
; Patent No. 6372453

GENERAL INFORMATION:

APPLICANT: Robert D. Klein
TITLE OF INVENTION: Neurturin Receptor

NUMBER OF SEQUENCES: 28

CORRESPONDENCE ADDRESS:

ADDRESSEE: Genentech, Inc.

STREET: 1 DNA Way

CITY: South San Francisco

STATE: California

COUNTRY: USA

ZIP: 94080

COMPUTER READABLE FORM:

MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: WinPatIn (Genentech)

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/802,805D

FILING DATE: 18-Feb-1997

CLASSIFICATION: 536

ATTORNEY/AGENT INFORMATION:

NAME: Torchia, PhD., Timothy E.

REGISTRATION NUMBER: 36,700

REFERENCE/DOCKET NUMBER: P1086

TELECOMMUNICATION INFORMATION:

TELEPHONE: 650/225-8674

TELEFAX: 650/952-9881

INFORMATION FOR SEQ ID NO: 3:

SEQUENCE CHARACTERISTICS:

LENGTH: 464 amino acids

TYPE: Amino Acid

TOPOLOGY: Linear

US-08-802-805D-3

Query Match 6.5%; Score 83; DB 4; Length 464;

Best Local Similarity 22.3%; Pred. No. 0.32;

Matches 60; Conservative 23; Mismatches 94; Indels 92; Gaps 13;

```
QY 6 VRADLVFLQGLAESREQAKRLIMAGK--VTLTNNSTTIPRL--EKP-----GHK 52
Db 41 VRANELCAEASNCSSRYRTLRQCLAGDRNTMLANKECQAALVQLQESPLYDCRCRGMK 100
QY 53 YPLESI-----CSLIGVERFVSRGAYKLLTA--LDFFK-----IDVKSCI 90
Db 101 KELQCLQIYWSIHLGTEGEFEYASPYEPTVSRSLDIFRLASIFSGTGADPVVSAKSNH 160
QY 91 CLDAGASTGGTDCLLQHGASKVYADVKGQOLHEKLYTNEQVINTGVNLRTASK--- 146
Db 161 CLDAKA-----CNLNDCKLRSSYIS-----ICNREISPTERCNRKCHALRQ 206
QY 147 --DLIPEVDILTIDVSFISL-----TLILPSCIRWLKASGIIIALIKPQFELYP 194
Db 207 FFDVFPSE---YTYRMLFCSQDQACAEERRRQTILPSC-----SYE 244
QY 195 DKIKKGWVKETSLQYEAWEKIIHFQCOSEL 223
Db 245 DK-----EKNCLDLRGVCRTHLCSRL 268
```

RESULT 24

US-09-187-906-13

; Sequence 13, Application US/09187906

; Patent No. 6677135

GENERAL INFORMATION:

APPLICANT: BIOGEN, INC.

TITLE OF INVENTION: Ret Ligand (RetL) for Stimulating Neural

TITLE OF INVENTION: and Renal Growth
NUMBER OF SEQUENCES: 21
CORRESPONDENCE ADDRESS:
ADDRESSEE: Biogen, Inc.
STREET: 14 Cambridge Center
CITY: Cambridge
STATE: MA
COUNTRY: USA
ZIP: 02142
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/187,906
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/US97/07726
FILING DATE: 07-MAY-97
APPLICATION NUMBER: US 60/017,427
FILING DATE: 08-MAY-96
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/019,300
FILING DATE: 07-JUN-96
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/021,859
FILING DATE: 16-JUL-96
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/043,533
FILING DATE: 10-APR-97
ATTORNEY/AGENT INFORMATION:
NAME: Kaplan, Warren A.
REGISTRATION NUMBER: 34,199
REFERENCE/DOCKET NUMBER: A008 PCT CIP
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-679-2400
TELEFAX: 617-679-2838
INFORMATION FOR SEQ ID NO: 13:
SEQUENCE CHARACTERISTICS:
LENGTH: 464 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-187-906-13

Query Match 6.5%; Score 83; DB 4; Length 464;

Best Local Similarity 22.3%; Pred. No. 0.32;

Matches 60; Conservative 23; Mismatches 94; Indels 92; Gaps 13;

```
QY 6 VRADLVFLQGLAESREQAKRLIMAGK--VTLTNNSTTIPRL--EKP-----GHK 52
Db 41 VRANELCAEASNCSSRYRTLRQCLAGDRNTMLANKECQAALVQLQESPLYDCRCRGMK 100
QY 53 YPLESI-----CSLIGVERFVSRGAYKLLTA--LDFFK-----IDVKSCI 90
Db 101 KELQCLQIYWSIHLGTEGEFEYASPYEPTVSRSLDIFRLASIFSGTGADPVVSAKSNH 160
QY 91 CLDAGASTGGTDCLLQHGASKVYADVKGQOLHEKLYTNEQVINTGVNLRTASK--- 146
Db 161 CLDAKA-----CNLNDCKLRSSYIS-----ICNREISPTERCNRKCHALRQ 206
QY 147 --DLIPEVDILTIDVSFISL-----TLILPSCIRWLKASGIIIALIKPQFELYP 194
Db 207 FFDVFPSE---YTYRMLFCSQDQACAEERRRQTILPSC-----SYE 244
QY 195 DKIKKGWVKETSLQYEAWEKIIHFQCOSEL 223
Db 245 DK-----EKNCLDLRGVCRTHLCSRL 268
```

RESULT 25

US-08-957-063-16
; Sequence 16, Application US/08957063
; Patent No. 6025157
; GENERAL INFORMATION:
; APPLICANT: Robert D. Klein, Arnon Rosenthal, Mary A. Hynes
; TITLE OF INVENTION: Neurturin Receptor
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 1 DNA Way
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WinPatIn (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/487,685
; FILING DATE: 19-Jan-2000
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/957,063
; FILING DATE: <Unknown>
; APPLICATION NUMBER: 913
; FILING DATE: 18-Feb-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Torchia, PhD., Timothy E.
; REGISTRATION NUMBER: 36,700
; REFERENCE/DOCKET NUMBER: P1086P2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650/225-9874
; TELEFAX: 650/952-9881
; INFORMATION FOR SEQ ID NO: 16:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 664 amino acids
; TYPE: Amino Acid
; TOPOLOGY: Linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 16:
US-09-487-685-16
Query Match 6.5%; Score 83; DB 4; Length 664;
Best Local Similarity 22.3%; Pred. No. 0.57; Indels 92; Gaps 13;
Matches 60; Conservative 23; Mismatches 94; Indels 92; Gaps 13;
QY 6 VRADELVFLQGLAESREQAQLIMAGK---VLTNNSTTIPLRL--EXP-----GHK 52
DB 41 VRANELCAESNCSSRYRTLQCLAGRDNTMLANKEQAALVLEQESPLYDCRCRGMK 100
QY 53 YPLESI-----CSLIGVERFVSRGAYKLTA--LDFPK-----IDVSCI 90
DB 101 KELQCLQIYWSIHLGLTEGEFEYEASPEYPTVTSRLSDIFRLASIFSGTGADPVSASNSH 160
QY 91 CLDAGASTGGFTDCLLQHGASKVVAIDVGKQLHEKLYTNEOVINIEGVNLTASK--- 146
DB 161 CLDAAKA-----CNLNDNCKLRSSYS-ICNREISPTERCNRKCKHAKLRQ 206
QY 147 --DLIPEEVDILTIDVPSFSL-----TLILPSCIRWLKASGIIIALIKPOFELYP 194
DB 207 FDRVPSE---YTYRMLFSCQDQACAEERRQTILPSC-----SYE 244
QY 195 DKIKKGVVKETSLOYEAVEKIIHFQCSL 223
DB 245 DK-----EKPNCLDLRGVCRTHLCSRSL 268
RESULT 27
US-08-802-805D-16
; Sequence 16, Application US/08802805D
; Patent No. 6372453
; GENERAL INFORMATION:
; APPLICANT: Robert D. Klein
; TITLE OF INVENTION: Neurturin Receptor
; NUMBER OF SEQUENCES: 28
; CORRESPONDENCE ADDRESS:

US-08-957-063-16
; Sequence 16, Application US/08957063
; Patent No. 6025157
; GENERAL INFORMATION:
; APPLICANT: Robert D. Klein, Arnon Rosenthal, Mary A. Hynes
; TITLE OF INVENTION: Neurturin Receptor
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 1 DNA Way
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WinPatIn (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/957,063
; FILING DATE: 24-Oct-1997
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/871
; FILING DATE: 9-Jun-1997
; APPLICATION NUMBER: 913
; FILING DATE: 18-Feb-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Torchia, PhD., Timothy E.
; REGISTRATION NUMBER: 36,700
; REFERENCE/DOCKET NUMBER: P1086P2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650/225-9874
; TELEFAX: 650/952-9881
; INFORMATION FOR SEQ ID NO: 16:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 664 amino acids
; TYPE: Amino Acid
; TOPOLOGY: Linear
US-08-957-063-16
Query Match 6.5%; Score 83; DB 3; Length 664;
Best Local Similarity 22.3%; Pred. No. 0.57; Indels 92; Gaps 13;
Matches 60; Conservative 23; Mismatches 94; Indels 92; Gaps 13;
QY 6 VRADELVFLQGLAESREQAQLIMAGK---VLTNNSTTIPLRL--EXP-----GHK 52
DB 41 VRANELCAESNCSSRYRTLQCLAGRDNTMLANKEQAALVLEQESPLYDCRCRGMK 100
QY 53 YPLESI-----CSLIGVERFVSRGAYKLTA--LDFPK-----IDVSCI 90
DB 101 KELQCLQIYWSIHLGLTEGEFEYEASPEYPTVTSRLSDIFRLASIFSGTGADPVSASNSH 160
QY 91 CLDAGASTGGFTDCLLQHGASKVVAIDVGKQLHEKLYTNEOVINIEGVNLTASK--- 146
DB 161 CLDAAKA-----CNLNDNCKLRSSYS-ICNREISPTERCNRKCKHAKLRQ 206
QY 147 --DLIPEEVDILTIDVPSFSL-----TLILPSCIRWLKASGIIIALIKPOFELYP 194
DB 207 FDRVPSE---YTYRMLFSCQDQACAEERRQTILPSC-----SYE 244
QY 195 DKIKKGVVKETSLOYEAVEKIIHFQCSL 223
DB 245 DK-----EKPNCLDLRGVCRTHLCSRSL 268
RESULT 26
US-09-487-685-16
; Sequence 16, Application US/09487685
; Patent No. 6342348
; GENERAL INFORMATION:

ADDRESSEE: Genentech, Inc.
STREET: 1 DNA Way
CITY: South San Francisco
STATE: California
COUNTRY: USA
ZIP: 94080
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: WinPatIn (Genentech)
CURRENT APPLICATION DATA: US/08/802.80SD
FILING DATE: 18-Feb-1997
CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:
NAME: Torchia, PhD., Timothy E.
REGISTRATION NUMBER: 36,700
REFERENCE/DOCKET NUMBER: P1086
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650/225-8674
TELEFAX: 650/952-9881
INFORMATION FOR SEQ ID NO: 16:
SEQUENCE CHARACTERISTICS:
LENGTH: 664 amino acids
TYPE: Amino Acid
TOPOLOGY: Linear
US-08-802-80SD-16

Query Match 6.5%; Score 83; DB 4; Length 664;
Best Local Similarity 22.3%; Pred. No. 0.57; Indels 94; Gaps 13;
Matches 60; Conservative 23; Mismatches 94; Indels 92; Gaps 13;
QY 6 VRADELVFLQGLAESREQAELIMAGK---VTLTNNSTTIPLRL--EKP-----GHK 52
Db 41 VRANELCAESNCSSRYTLRQCLAGDRNTMLANKECQALEVLQSPHYDCRCRGMK 100
QY 53 YPLESI-----CSLIGVRFVSRGAYKLLTA--LDFFK-----IDVKSCL 90
Db 101 KELQCLQIYWSIHGLTGEFEYASPEYPTSRSLDIFRLASTFSCTGADPVVSASNSH 160
QY 91 CLDAGASTGGFTCLLQHGASKVVAIDVGKQLHEKLYTNEQVINIEGVNLTASK----- 146
Db 161 CLDAKA-----CNLNDNCKLRSSYIS-----ICNREISPTERCNRRKHKALRQ 206
QY 147 --DLIPEEVDILAIQVDFISL-----TLIPSCIRWLKASGIIIALIKPQFELYP 194
Db 207 FFDVFPSE---YTVMLFCSCQDQACAEPRRQTLIPSC-----SYE 244
QY 195 DKIKKGVKETSLOYEAVEKIIHFCQSEL 223
Db 245 DK-----EKNPCLDLGRVCRTDHLCSRL 268

RESULT 28
US-09-919-172-29
; Sequence 29, Application US/09919172
; Patent No. 6673545
; GENERAL INFORMATION:
; APPLICANT: Faris, Mary
; APPLICANT: Turner, Christopher M.
; TITLE OF INVENTION: PROSTATE CANCER MARKERS
; FILE REFERENCE: PA-0036 US
; CURRENT APPLICATION NUMBER: US/09/919,172
; CURRENT FILING DATE: 2001-07-30
; PRIOR APPLICATION NUMBER: 60/222,469
; PRIOR FILING DATE: 2000-07-28
; NUMBER OF SEQ ID NOS: 102
; SOFTWARE: PERL Program
; SEQ ID NO 29
; LENGTH: 2125
; TYPE: PRT
; ORGANISM: Homo sapiens

FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. 6673545 3774181CD1
US-09-919-172-29

Query Match 6.5%; Score 83; DB 4; Length 2125;
Best Local Similarity 24.4%; Pred. No. 3.7; Indels 52; Gaps 9;
Matches 53; Conservative 29; Mismatches 83; Indels 52; Gaps 9;
QY 31 GKVTLTNNSTTIPLR-----LEKPGHKYPLESTCSLIGVRFVSRGAYKLLTALDPFKI 84
Db 1899 GLITLTSLADSLSLRLVPKDLSPVAGYWLTAGSERISVLKASRRNLVDRIITALR----- 1954
QY 85 DVKSCICLDAGASTGGFTCDLQHGASKVY-----AIDVGKQLHEKL----- 127
Db 1955 -----CLEAQVSTGGIIDEI-----TGKTRVAFALHRLGVDEGFAQLQCELVITGIG 2004
QY 128 --YTNEQVINIEGVNLTASKDLIPERVDILTIDVFSITLTILPSCIRWLKAS-----G 180
Db 2005 HPITNKMSVVEAVN-----ANIINKEMGIRCLFQYLTGGLIEPQVHSRLSIEEALQVG 2059
QY 181 IIALIKPQFELYPDKIKGVVKET--SLOY-EAVEK 214
Db 2060 IIDVLATKLDKQKSYVRNIICPTQKRLTYKSALEK 2096

RESULT 29

US-09-328-352-5334
; Sequence 5334, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
; FILE REFERENCE: GTC99-03PA
; CURRENT APPLICATION NUMBER: US/09/328,352
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 5334
; LENGTH: 340
; TYPE: PRT
; ORGANISM: Acinetobacter baumannii
US-09-328-352-5334

Query Match 6.5%; Score 82.5; DB 4; Length 340;
Best Local Similarity 24.1%; Pred. No. 0.22; Indels 67; Gaps 14;
Matches 62; Conservative 36; Mismatches 92; Indels 67; Gaps 14;
QY 6 VRADELVFLQGLAES--REQAKELIMAGKVTLTNNSTTIP---LRLEKPGHKYPLES--- 57
Db 98 LKTDQSVFLVGEKKGVVERAAKQLQGFGLKILKDSARHCQLWHLKIEKIKPLESWLK 157
QY 58 -----ICSLIGV--ERFVSRGAYKLLTALDFFKIDVKSCICLDAGASTGGFTD 103
Db 158 TYTVQVNEQELTICALPGVFSQTHLDVGTAVLLPYLN---QVKSRIADFGCG-AGIIS 212
QY 104 CLIQHGASK--VVAIDVGKQLHEKLYTNEQVINIEGV---NLR-----TASKDLIPEVD 154
Db 213 CYLAKANSSNIHALDIDAFALQ-----STEMFSGNGIGSDQLRLOPVTGIAD-APELD 267
QY 155 ILT-----IDVFSITLTILPSCIRWLKASGIIIALIKPQFELYPDKIKGVVKETS 206
Db 268 AIVSNPPFHQGIHTNYDASEGLCQNAKHLKASG-ELWIVANRFLNYPILIEK----- 319
QY 207 LQYEAKEKIIHRCQSEL 223
Db 320 -----HFGQCEI 326

RESULT 30
US-08-957-063-6
; Sequence 6, Application US/08957063
; Patent No. 6025157

; GENERAL INFORMATION:
 ; APPLICANT: Robert D. Klein, Arnon Rosenthal, Mary A. Hynes
 ; TITLE OF INVENTION: Neurturin Receptor
 ; NUMBER OF SEQUENCES: 19
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Genentech, Inc.
 ; STREET: 1 DNA Way
 ; CITY: South San Francisco
 ; STATE: California
 ; COUNTRY: USA
 ; ZIP: 94080
 ;
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: WinPatIn (Genentech)
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/957,063
 ; FILING DATE: 24-Oct-1997
 ; CLASSIFICATION: 800
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: 08/871
 ; FILING DATE: 9-Jun-1997
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: 913
 ; FILING DATE: 18-Feb-1997
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Torchia, P.D., Timothy E.
 ; REGISTRATION NUMBER: 36,700
 ; REFERENCE/DOCKET NUMBER: P1086P2
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 650/225-8674
 ; TELEFAX: 650/952-9881
 ; INFORMATION FOR SEQ ID NO: 6:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 464 amino acids
 ; TYPE: Amino Acid
 ; TOPOLOGY: Linear
 ;
 ; US-08-957-063-6

Query Match 6.4%; Score 81; DB 3; Length 464;
 Best Local Similarity 21.9%; Pred. No. 0.57;
 Matches 59; Conservative 25; Mismatches 93; Indels 92; Gaps 13;

 QY 6 VRADLVFLQGLAESREQAKRLIMAGK---VLTNNSTIPLRL--EKP-----GHK 52
 DB 41 VRANELCAAESNCSSRYTLRQCLAGDRNTWLANKECOALEVLOESPLYDCRCRGMK 100
 QY 53 YPLESI-----CSLIGVERFVSRGAYKLLTA--LDFFK-----IDVKSCI 90
 DB 101 KELQCLQIYWSHGLGTEGEFEYASPEVTSRLSDIFRLASIFSGTGTDPAVSTKSNH 160
 QY 91 CLDAGASTGGFTDCLLQHGASKVYADVKGQOLHEKLYTNEQVINIEGVNLTASK---- 146
 DB 161 CLDAKA-----CNLNDNCKLRSSYIS-----ICNREISPTERCNRRKCHKALRQ 206
 QY 147 --DLIPEEVDILTIDVSFISL-----TLILPSCIRMLKASGIIIALIKPQFELYP 194
 DB 207 FFDVRPSE---YTYRMLFCSCDQQAERRRQTILPSC-----SYE 244

 QY 195 DKIKKGWVKETSLQVEAVEKEIHFQOSEL 223
 DB 245 DK-----EKPNCLDLRLSLCRTDHLCSRL 268

 RESULT 31
 US-09-487-685-6
 ; Sequence 6, Application US/09487685
 ; Patent No. 6342348
 ; GENERAL INFORMATION:
 ; APPLICANT: Robert D. Klein, Arnon Rosenthal, Mary A. Hynes
 ; TITLE OF INVENTION: Neurturin Receptor
 ; NUMBER OF SEQUENCES: 19

; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Genentech, Inc.
 ; STREET: 1 DNA Way
 ; CITY: South San Francisco
 ; STATE: California
 ; COUNTRY: USA
 ; ZIP: 94080
 ;
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: WinPatIn (Genentech)
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/09/487,685
 ; FILING DATE: 19-Jan-2000
 ; CLASSIFICATION: <Unknown>
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: 08/957,063
 ; FILING DATE: <Unknown>
 ; APPLICATION NUMBER: 913
 ; FILING DATE: 18-Feb-1997
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Torchia, P.D., Timothy E.
 ; REGISTRATION NUMBER: 36,700
 ; REFERENCE/DOCKET NUMBER: P1086P2
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 650/225-8674
 ; TELEFAX: 650/952-9881
 ; INFORMATION FOR SEQ ID NO: 6:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 464 amino acids
 ; TYPE: Amino Acid
 ; TOPOLOGY: Linear
 ;
 ; SEQUENCE DESCRIPTION: SEQ ID NO: 6:
 ;
 ; US-09-487-685-6

 Query Match 6.4%; Score 81; DB 4; Length 464;
 Best Local Similarity 21.9%; Pred. No. 0.57;
 Matches 59; Conservative 25; Mismatches 93; Indels 92; Gaps 13;

 QY 6 VRADLVFLQGLAESREQAKRLIMAGK---VLTNNSTIPLRL--EKP-----GHK 52
 DB 41 VRANELCAAESNCSSRYTLRQCLAGDRNTWLANKECOALEVLOESPLYDCRCRGMK 100
 QY 53 YPLESI-----CSLIGVERFVSRGAYKLLTA--LDFFK-----IDVKSCI 90
 DB 101 KELQCLQIYWSHGLGTEGEFEYASPEVTSRLSDIFRLASIFSGTGTDPAVSTKSNH 160
 QY 91 CLDAGASTGGFTDCLLQHGASKVYADVKGQOLHEKLYTNEQVINIEGVNLTASK---- 146
 DB 161 CLDAKA-----CNLNDNCKLRSSYIS-----ICNREISPTERCNRRKCHKALRQ 206
 QY 147 --DLIPEEVDILTIDVSFISL-----TLILPSCIRMLKASGIIIALIKPQFELYP 194
 DB 207 FFDVRPSE---YTYRMLFCSCDQQAERRRQTILPSC-----SYE 244

 QY 195 DKIKKGWVKETSLQVEAVEKEIHFQOSEL 223
 DB 245 DK-----EKPNCLDLRLSLCRTDHLCSRL 268

 RESULT 32
 US-08-802-805D-6
 ; Sequence 6, Application US/08802805D
 ; Patent No. 6372453
 ; GENERAL INFORMATION:
 ; APPLICANT: Robert D. Klein
 ; TITLE OF INVENTION: Neurturin Receptor
 ; NUMBER OF SEQUENCES: 28
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Genentech, Inc.
 ; STREET: 1 DNA Way
 ; CITY: South San Francisco

```

; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WinPatIn (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/802.805D
; FILING DATE: 18-Feb-1997
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Torchia, PhD., Timothy E.
; REGISTRATION NUMBER: 36,700
; REFERENCE/DOCKET NUMBER: P1086
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650/225-8674
; TELEFAX: 650/952-9881
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 464 amino acids
; TYPE: Amino Acid
; TOPOLOGY: Linear
; US-08-802-805D-6

```

```

Query Match
Best Local Similarity 6.4%; Score 81; DB 4; Length 464;
Matches 59; Conservative 25; Mismatches 93; Indels 92; Gaps 13;

QY 6 VRDELVFLOGLAESRQAKRLIMAGK---VTLTNNSTTIPLRL--EKP-----GHK 52
Db 41 VRANELCAESNCSSRYRTLQCLAGDRNTMLANKCQAALVLOESPLYDCRCKRGMK 100
QY 53 YPLESI-----CSLIGVERFVSRGAYKLLTA--LDFFK-----IDVKSCI 90
Db 101 KELQCLQIYWSIHGLTEGEFEYASPEVPTSLSDFIRLASIFSGTGDPVSTKSNH 160
QY 91 CLDAGSTGGTDCLLQHGASKVYADVKGQLEKLYTNEQVINIEGVNLRTASK--- 146
Db 161 CLDAKA-----CNLNDNCKLRSSYIS-----ICNREISPTERCNRKCHKALRQ 206
QY 147 --DLIPEVDILTIDVSFISL-----TLILPSCIWLKASGIIIALIKPQFELYP 194
Db 207 FFRVPSSE---YYRMLFCSQDQACARRRQTLPLSC-----SYE 244
QY 195 DKIKKGVVKETSQYEAWEKIIHFQCSSEL 223
Db 245 DK-----EKNPCLDLRSLCRTDHLCSRL 268

```

```

RESULT 33
US-08-957-063-18
; Sequence 18, Application US/08957063
; Patent No. 6025157
; GENERAL INFORMATION:
; APPLICANT: Robert D. Klein, Arnon Rosenthal, Mary A. Hynes
; TITLE OF INVENTION: Neurturin Receptor
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 1 DNA Way
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WinPatIn (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/957,063

```

```

; FILING DATE: 24-Oct-1997
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/871
; FILING DATE: 9-Jun-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 913
; FILING DATE: 18-Feb-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Torchia, PhD., Timothy E.
; REGISTRATION NUMBER: 36,700
; REFERENCE/DOCKET NUMBER: P1086P2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650/225-8674
; TELEFAX: 650/952-9881
; INFORMATION FOR SEQ ID NO: 18:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 664 amino acids
; TYPE: Amino Acid
; TOPOLOGY: Linear
; US-08-957-063-18

```

```

Query Match
Best Local Similarity 6.4%; Score 81; DB 3; Length 664;
Matches 59; Conservative 25; Mismatches 93; Indels 92; Gaps 13;

QY 6 VRDELVFLOGLAESRQAKRLIMAGK---VTLTNNSTTIPLRL--EKP-----GHK 52
Db 41 VRANELCAESNCSSRYRTLQCLAGDRNTMLANKCQAALVLOESPLYDCRCKRGMK 100
QY 53 YPLESI-----CSLIGVERFVSRGAYKLLTA--LDFFK-----IDVKSCI 90
Db 101 KELQCLQIYWSIHGLTEGEFEYASPEVPTSLSDFIRLASIFSGTGDPVSTKSNH 160
QY 91 CLDAGSTGGTDCLLQHGASKVYADVKGQLEKLYTNEQVINIEGVNLRTASK--- 146
Db 161 CLDAKA-----CNLNDNCKLRSSYIS-----ICNREISPTERCNRKCHKALRQ 206
QY 147 --DLIPEVDILTIDVSFISL-----TLILPSCIWLKASGIIIALIKPQFELYP 194
Db 207 FFRVPSSE---YYRMLFCSQDQACARRRQTLPLSC-----SYE 244
QY 195 DKIKKGVVKETSQYEAWEKIIHFQCSSEL 223
Db 245 DK-----EKNPCLDLRSLCRTDHLCSRL 268

```

```

RESULT 34
US-09-487-685-18
; Sequence 18, Application US/09487685
; Patent No. 6342348
; GENERAL INFORMATION:
; APPLICANT: Robert D. Klein, Arnon Rosenthal, Mary A. Hynes
; TITLE OF INVENTION: Neurturin Receptor
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 1 DNA Way
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WinPatIn (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/487,685
; FILING DATE: 19-Jan-2000
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/957,063

```

```

; FILING DATE: <Unknown>
; APPLICATION NUMBER: 913
; FILING DATE: 18-Feb-1997
; ATTORNEY/AGENT INFORMATION:
;   NAME: Torchia, PhD., Timothy E.
;   REGISTRATION NUMBER: 36,700
;   REFERENCE/DOCKET NUMBER: P1086P2
; TELECOMMUNICATION INFORMATION:
;   TELEPHONE: 650/225-8674
;   TELEFAX: 650/952-9881
; INFORMATION FOR SEQ ID NO: 18:
;   SEQUENCE CHARACTERISTICS:
;     LENGTH: 664 amino acids
;     TYPE: Amino Acid
;     TOPOLOGY: Linear
;   SEQUENCE DESCRIPTION: SEQ ID NO: 18:
US-09-487-685-18

Query Match          6.4%; Score 81; DB 4; Length 664;
Best Local Similarity 21.9%; Pred. No. 1;
Matches 59; Conservative 25; Mismatches 93; Indels 92; Gaps 13;

QY 6 VRADLVFLOGLABREQAKLIMAGK---VTLNNSTTIPRL--EKP-----GHK 52
DB 41 VRANELCAESNCSSRYRTLRLQCLAGRDNTMLANKECQAALVQLQSPDYDCRCRGMK 100
QY 53 YPLESI-----CSLIGVERFVSRGAYKLLTA--LDFFK-----IDVKSCI 90
DB 101 KELQCLQIYWSIHGLTGEFEFYEASPEYPTSLRSLDFRLASIFSGTGTDPAVSTKSNH 160
QY 91 CLDAGASTGGTDCCLLQHGASKVVAIDVGKQLHEKLYTNEQVINIEGVNLRTASK---- 146
DB 161 CLDRAKA-----CNLNDNCKLRSSYIS-----ICNREISPTERCNRKCHKALRQ 206
QY 147 --DLIPEVDILTIDVSPISL-----TLILPSCIRWLKASGIIILIKPOFELYP 194
DB 207 FFDVRPSE---YTYRMLFCSQDQACAEARRRQTILPSC-----SYE 244
QY 195 DKIKGVVVKETSLQYEAKEIHFQSEL 223
DB 245 DK-----EKNCLDLRLSLCRTDHLCSRL 268

```

```

RESULT 35
US-08-802-805D-18
; Sequence 18, Application US/08802805D
; Patent No. 6372453
; GENERAL INFORMATION:
; APPLICANT: Robert D. Klein
; TITLE OF INVENTION: Neurturin Receptor
; NUMBER OF SEQUENCES: 28
; CORRESPONDENCE ADDRESS:
; ADDRESS: Genentech, Inc.
; STREET: 1 DNA Way
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WinPatIn (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/802,805D
; FILING DATE: 18-Feb-1997
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Torchia, PhD., Timothy E.
; REGISTRATION NUMBER: 36,700
; REFERENCE/DOCKET NUMBER: P1086
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650/225-8674

```

```

; TELEFAX: 650/952-9881
; INFORMATION FOR SEQ ID NO: 18:
; SEQUENCE CHARACTERISTICS:
;   LENGTH: 664 amino acids
;   TYPE: Amino Acid
;   TOPOLOGY: Linear
; US-08-802-805D-18

Query Match          6.4%; Score 81; DB 4; Length 664;
Best Local Similarity 21.9%; Pred. No. 1;
Matches 59; Conservative 25; Mismatches 93; Indels 92; Gaps 13;

QY 6 VRADLVFLOGLABREQAKLIMAGK---VTLNNSTTIPRL--EKP-----GHK 52
DB 41 VRANELCAESNCSSRYRTLRLQCLAGRDNTMLANKECQAALVQLQSPDYDCRCRGMK 100
QY 53 YPLESI-----CSLIGVERFVSRGAYKLLTA--LDFFK-----IDVKSCI 90
DB 101 KELQCLQIYWSIHGLTGEFEFYEASPEYPTSLRSLDFRLASIFSGTGTDPAVSTKSNH 160
QY 91 CLDAGASTGGTDCCLLQHGASKVVAIDVGKQLHEKLYTNEQVINIEGVNLRTASK---- 146
DB 161 CLDRAKA-----CNLNDNCKLRSSYIS-----ICNREISPTERCNRKCHKALRQ 206
QY 147 --DLIPEVDILTIDVSPISL-----TLILPSCIRWLKASGIIILIKPOFELYP 194
DB 207 FFDVRPSE---YTYRMLFCSQDQACAEARRRQTILPSC-----SYE 244
QY 195 DKIKGVVVKETSLQYEAKEIHFQSEL 223
DB 245 DK-----EKNCLDLRLSLCRTDHLCSRL 268

```

```

RESULT 36
US-09-134-001C-3570
; Sequence 3570, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; FILE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 3570
; LENGTH: 406
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
; US-09-134-001C-3570

Query Match          6.3%; Score 80.5; DB 4; Length 406;
Best Local Similarity 19.2%; Pred. No. 0.53;
Matches 45; Conservative 45; Mismatches 67; Indels 77; Gaps 11;

QY 42 IPLRLKPGHKYPLESICSLIGVERFVSRGAYKLLTALDFFKIDVKSCICLDAGASTGGF 101
DB 2 IELPIERNGHMLSKNLKSK-----VYAGGKK---AVDNMTIDIES-----GDF 41
QY 102 TDCLLQHGASKVVAIDVGKQLHEKLYTNEQVINIEGVNLRATASKDLIPEE----- 152
DB 42 IAFITGSGGKTALRM-----INRMISTEGEITIDGKNI-----KELNPVLRRSIGVIV 93
QY 153 -----VDILTIDVSPISLTLILPSCIRWL-----KASGII--IALIKPOFELYDPKI-- 197
DB 94 QOIGLPHMTYKENIV-----LVPKLLKWSQEKOKAKELRLVLDLPEYLDRYSELSG 149
QY 198 ----KKGVVVKETSLQYEA-----EKIHFQSELGLIFIGV 230

```

Db 150 GQQRIGVVRALAAEQDIIIMDEPFGLADPTITROTLDLVKKLQQQIGKTFIFV 203

RESULT 37

US-09-328-352-6056

; Sequence 6056, Application US/09328352

; Patent No. 6562958

; GENERAL INFORMATION:

; APPLICANT: Gary L. Breton et al.

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER

; FILE REFERENCE: GTC99-03PA

; CURRENT APPLICATION NUMBER: US/09/328,352

; CURRENT FILING DATE: 1999-06-04

; NUMBER OF SEQ ID NOS: 8252

; SEQ ID NO 6056

; LENGTH: 276

; TYPE: PRT

; ORGANISM: Acinetobacter baumannii

US-09-328-352-6056

Query Match 6.3%; Score 80; DB 4; Length 276;

Best Local Similarity 25.2%; Pred. No. 0.33;

Matches 37; Conservative 26; Mismatches 72; Indels 12; Gaps 5;

QY 51 HKY----PLESICSLIGVERFVSRGAYK-LLTALDFFKIDVKSCICLDAGASTGGFTDCLL 106

Db 47 HKHSKNPDPESVDARRAFLQGGYQPLQAVVHLLKQKAKMVLIDGCGEGYTSA-M 105

QY 107 QHGASKVYVIDVGKQGLHEKLYTNEQVINIEGVNLTASKDLPIEEVDILTIDVSFISLT 166

Db 106 QQVVEQIGVDIAKNVQRAUKLVTVVVGTTG---ATLPIVDQSDMDVCTSLFSPITQ 162

QY 167 LILPSCIRWLKASGIIITALIKPOFELY 193

Db 163 EIL----RVLKDDGYLIVTPATDHLV 185

RESULT 38

US-09-540-236-2477

; Sequence 2477, Application US/09540236

; Patent No. 6673910

; GENERAL INFORMATION:

; APPLICANT: Gary L. Breton et al.

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO MORAXELLA CATAR

; FILE REFERENCE: 2709.2005-001

; CURRENT APPLICATION NUMBER: US/09/540,236

; CURRENT FILING DATE: 2000-04-04

; NUMBER OF SEQ ID NOS: 3840

; SEQ ID NO 2477

; LENGTH: 287

; TYPE: PRT

; ORGANISM: M.catarrhalis

US-09-540-236-2477

Query Match 6.3%; Score 80; DB 4; Length 287;

Best Local Similarity 22.6%; Pred. No. 0.35;

Matches 45; Conservative 40; Mismatches 80; Indels 34; Gaps 9;

QY 17 LAESRQAKELINAGKVTLTNNSTTIP-LRLEKPGHKYPLESICSLIGVERFVSRGAYKL 75

Db 47 IADAGDEAE--KAGGVKLTANQRIIPFMKLEKPMIVFVLSI--LIGLTAIVKG---L 99

QY 76 LTALDF-----FKIDVKSCICLDAGASTGGFTDCLLQHGASK---VVAIDVGKQL 123

Db 100 NFGLDTGGVSANVAYEQVEQAQVIKALAD-NGENDAVQYLGTRQELLVRLPQDGS 158

QY 124 HEKLYTNEQVINIEGVNLTASKDLPIEEVDILTIDVSFISLTILIPSCIRWLKASGII 183

Db 159 ENLSASLDQILNFPN-----NPSITDSISIGSQLGNEIYLSLAMSVALGMML 208

QY 184 ALIKPOFELYPDKIKKGW 202

Db 209 IYVATRFQF---KLALGAV 224

RESULT 39

US-08-472-534-5

; Sequence 5, Application US/08472534

; Patent No. 5919620

; GENERAL INFORMATION:

; APPLICANT: Hamel, Josee

; APPLICANT: Brodeur, Bernard R

; APPLICANT: Martin, Denis

; TITLE OF INVENTION: HEAT SHOCK PROTEIN HSP72 FROM

; NUMBER OF SEQUENCES: 6

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Fish & Neave

; STREET: 1251 Avenue of the Americas

; CITY: New York

; STATE: New York

; COUNTRY: United States of America

; ZIP: 10020

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/472,534

; FILING DATE:

; CLASSIFICATION: 424

; ATTORNEY/AGENT INFORMATION:

; NAME: Haley Jr, James F

; REGISTRATION NUMBER: 27,794

; REFERENCE/DOCKET NUMBER: Biovac-2

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 212-596-9000

; TELEFAX: 212-596-9090

; TELEX: 14-8367

; INFORMATION FOR SEQ ID NO: 5:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 607 amino acids

; TYPE: amino acid

; TOPOLOGY: linear

; MOLECULE TYPE: protein

US-08-472-534-5

Query Match 6.3%; Score 80; DB 2; Length 607;

Best Local Similarity 21.9%; Pred. No. 1.2;

Matches 54; Conservative 40; Mismatches 93; Indels 60; Gaps 11;

QY 14 LQGLAESRQAKELINAGKVTLTNNSTTIPLRLEKPGHKYPLESICSLIGVERFVSRGAY 73

Db 229 MQRLKDAEAKAKDLSG--VT-----STQSLPITAGEAGPLH-----LEWTLTRAKF 275

QY 74 KLLT--ALDPFKIDVKSCICLDAGASTGGFTDCLLQHGASKVYVIDVGKQGLHEKLYTNE 131

Db 276 DDLTRDLVERTKVPVQALS-DAGLSLSEIDEVILVGGSTRIPAV----- 319

QY 132 QVINIEGVNLTAA---SKDLIPEV-----DILTIDVSFISLTILIPSCIRWLKAS 179

Db 320 ----VEAVKAETGKPNKSVNPDVVAMGAALGGVITGDKVDVLLDVTPLSLGIETMG 375

QY 180 GTIIALIKPOFELYPDKIKKGWVKETSLQVEAVEKIIHFQCS-----LGLIFIGV 230

Db 376 GVFTKLIDNTTPTTSKQ--VFSTAADNQPAVD--IHVLQGRPMAADNKTILGRFLTD 431

QY 231 VPSVIKG 237

Db 432 IPAAPRG 438

RESULT 40

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 7, 2004, 16:56:51 ; Search time 70.8804 Seconds

(without alignments)
1102.311 Million cell updates/sec

Title: US-10-009-919A-1

Perfect score: 1271

Sequence: 1 MAKHKVRADELVLFLQGLAES.....PSVIKPGKNGEYLIVLKRR 251

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1276540 seqs, 311283816 residues

Total number of hits satisfying chosen parameters: 1276540

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:*

1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep.*
7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep.*
10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep.*
11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep.*
12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep.*
13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep.*
14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep.*
15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep.*
16: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep.*
17: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep.*
18: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	446.5	35.1	270	9	US-09-765-272-64
2	433.5	34.1	206	15	US-10-264-213-162
3	409.5	32.2	306	12	US-10-425-114-64526
4	385	30.3	312	12	US-10-424-599-235806
5	353	27.8	176	12	US-10-425-114-49907
6	347.5	27.3	273	9	US-09-738-626-5061
7	343.5	27.0	271	14	US-10-156-761-14023
8	307.5	24.2	235	12	US-10-335-977-7740
9	307.5	24.2	235	12	US-10-335-977-7741
10	307.5	24.2	246	12	US-10-335-977-7742
11	290.5	22.9	235	10	US-09-882-227-246
12	240	18.9	161	16	US-10-437-963-128703
13	198	15.6	131	12	US-10-425-114-56226
14	101.5	8.0	374	12	US-10-424-599-183840
15	93.5	7.4	642	12	US-10-282-122A-65357

93.5	7.4	642	12	US-10-282-122A-65784	Sequence 65784, A
93.5	7.4	642	16	US-10-389-566-1039	Sequence 1039, Ap
93	7.3	615	12	US-10-282-122A-53012	Sequence 53012, A
93	7.3	725	16	US-10-437-963-150375	Sequence 150375, A
92.5	7.3	444	12	US-10-282-122A-60355	Sequence 60355, A
91.5	7.2	399	12	US-10-072-012-554	Sequence 554, App
91.5	7.2	409	12	US-10-072-012-555	Sequence 555, App
89.5	7.0	628	15	US-10-369-493-10300	Sequence 10300, A
87.5	6.9	415	14	US-10-269-557-12	Sequence 12, Appl
87.5	6.9	494	12	US-10-282-122A-66894	Sequence 66894, A
87.5	6.9	562	14	US-10-269-557-14	Sequence 14, Appl
87.5	6.9	642	14	US-10-269-557-13	Sequence 13, Appl
87.5	6.9	642	14	US-10-269-557-15	Sequence 15, Appl
87.5	6.9	642	16	US-10-389-566-1232	Sequence 1292, Ap
87.5	6.9	662	14	US-10-269-557-16	Sequence 16, Appl
87.5	6.9	959	16	US-10-437-963-103572	Sequence 103572, A
87	6.8	131	12	US-10-282-122A-53447	Sequence 53447, A
87	6.8	313	12	US-10-424-599-159278	Sequence 159278, A
87	6.8	317	12	US-10-282-122A-72185	Sequence 72185, A
87	6.8	644	15	US-10-369-493-19647	Sequence 19647, A
86	6.8	1855	14	US-10-177-293-315	Sequence 315, App
85.5	6.7	221	12	US-10-424-599-144586	Sequence 144586, A
85.5	6.7	337	12	US-10-282-122A-45182	Sequence 45182, A
84.5	6.6	160	9	US-09-752-385-12	Sequence 12, Appl
84	6.6	203	12	US-10-282-122A-72484	Sequence 72484, A
84	6.6	349	12	US-10-282-122A-61836	Sequence 61836, A
84	6.6	739	12	US-10-282-122A-56964	Sequence 56964, A
83.5	6.6	272	12	US-10-282-122A-67355	Sequence 67355, A
83.5	6.6	380	16	US-10-437-963-133230	Sequence 133230, A
83.5	6.6	549	14	US-10-156-761-9309	Sequence 9309, Ap

ALIGNMENTS

RESULT 1

US-09-765-272-64

; Sequence 64, Application US/09765272

; Patent No. US20020061545A1

; GENERAL INFORMATION:

; APPLICANT: Choi et. al.

; TITLE OF INVENTION: Streptococcus pneumoniae Antigenes and Vaccines

; NUMBER OF SEQUENCES: 452

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Human Genome Sciences, Inc.

; STREET: 9410 Key West Avenue

; CITY: Rockville

; STATE: Maryland

; COUNTRY: USA

; ZIP: 20850

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage

; COMPUTER: HP Vectra 486/33

; OPERATING SYSTEM: MSDOS version 6.2

; SOFTWARE: ASCII Text

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/765,272

; FILING DATE: 22-Jan-2001

; CLASSIFICATION: <Unknown>

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 08/961,083

; FILING DATE: <Unknown>

; ATTORNEY/AGENT INFORMATION:

; NAME: Brookes, A. Anders

; REGISTRATION NUMBER: 36,373

; REFERENCE/DOCKET NUMBER: PB340P2

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (301) 309-8504

; TELEFAX: (301) 309-8512

; INFORMATION FOR SEQ ID NO: 64:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 270 amino acids

; TYPE: amino acid

```
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 64:
US-09-765-272-64

Query Match      35.1%; Score 446.5; DB 9; Length 270;
Best Local Similarity 42.7%; Pred. No. 5.9e-41;
Matches 106; Conservative 41; Mismatches 94; Indels 7; Gaps 4;

QY 5 KVRADLVFLQGLAESRQAKRLIMAGKVTLTNNSTTIPRLKPKGHKYPLESICSLIGV 64
Db 1 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
2 KERVVDLAYKQGLFETREQAKRGWAGVAVLNGE----RFDKPGKIPDPTDLKUKGE 57
QY 65 E-RPVSGAYKLLTALDPFKIDVKSCICLDAGASTGGFTDCLLQHGASKVYVIDVGKQL 123
Db 58 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
58 KLVYSRGGGLKLEKALQVFLSDVGATTIDIGASTGGFTDWMQLNSAKLVFAVDVGNQL 117
QY 124 HEKLYTNEQVINTGVNLRTASKDLIEEVDILTIDVSFISLTILPSCIRWLKASGII 183
Db 118 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
118 AWKLQDPVRVMEQFNRFAEKTDFEQEPSFASIDVSFISLTILPALHRLVADQGVV 177
QY 184 ALIKPQFELYDPDKI-KKGWVKETSLQYEAVEKIIHFQCSBELGLIFIGWPSVIKGPKNQ 242
Db 178 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
178 ALVKPQFEAGREQIKGNGIIRDAKVNQVLESVTAMA-VEVGSRVGLDPSPIQGGHNI 236
QY 243 EYLIYLYK 250
Db 237 EFLAYLYK 244

RESULT 2
US-10-264-213-162
; Sequence 162, Application US/10264213
; Publication No. US20040009490A1
; GENERAL INFORMATION:
; APPLICANT: Glenn, Matthew
; APPLICANT: Havukkala, Ilkka J
; APPLICANT: Lubbers, Mark William
; APPLICANT: Dekker, James
; TITLE OF INVENTION: Polynucleotides, materials incorporating
; FILE REFERENCE: them, and methods for using them.
; CURRENT APPLICATION NUMBER: US/10/264,213
; CURRENT FILING DATE: 2002-10-03
; NUMBER OF SEQ ID NOS: 253
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 162
; LENGTH: 276
; TYPE: PRT
; ORGANISM: Lactobacillus rhamnosus
US-10-264-213-162

Query Match      34.1%; Score 433.5; DB 15; Length 276;
Best Local Similarity 45.2%; Pred. No. 1.7e-39;
Matches 112; Conservative 35; Mismatches 92; Indels 9; Gaps 5;

QY 5 KVRADLVFLQGLAESRQAKRLIMAGKVTLTNNSTTIPRLKPKGHKYPLESICSLIGV 64
Db 6 KERVVDLLVEQGLFESREQAKRSWAGEVVDQNNQ-----RLDKPGVKIPGDAILHVKGK 60
QY 65 ER-FVSRGAYKLLTALDPFKIDVKSCICLDAGASTGGFTDCLLQHGASKVYVIDVGKQL 123
Db 61 KMPYVSRRGGLKAKALVKFAIDVTGKTVLDIGASTGGFTDVALQNGAKMSYALDVGNQL 120
QY 124 HEKLYTNEQVINTGVNLR-TASKDLIPEEVDILTIDVSFISLTILPSCIRWLKASGII 182
Db 121 ANKLREDPVKWERNVFRYSKPEDETDGPFVFMATDVFSIKLILPLKAILMHPGHA 180
QY 183 IALIKPQFELYDPDKI-KKGWVKETSLQYEAIVEKIIHFQCSBELGLIFIGWPSVIKGPKN 241
Db 181 VALIKPQFEAGFAHVGHKGVIRDPQVRDVLKMINFA-SATGYNVGLGLAYSPIKGGEGN 239
```

```
QY 242 QEVLIYLYK 249
Db 240 IEFLIHQ 247

RESULT 3
US-10-425-114-64526
; Sequence 64526, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 64526
; LENGTH: 306
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: LIB3069-025-C8_FLI.pap
US-10-425-114-64526

Query Match      32.2%; Score 409.5; DB 12; Length 306;
Best Local Similarity 38.1%; Pred. No. 9.6e-37;
Matches 98; Conservative 47; Mismatches 81; Indels 31; Gaps 5;

QY 1 MAKHKYRDELVLQGLAESRQAKRLIMAGKV-----TLTNNSTTIPRLKPK 49
Db 47 LPKKRRLDEVCLERFQQYSRTYIQSWILQGVIVDGRVNVNKGATQVDSKSVIEIKAEIP 106
QY 50 GHKYPLESICSLIGVERFVSRGAYKLITALDFFKIDVKSCICLDAGASTGGFTDCLLQHG 109
Db 107 -----KYVCRAGHLEAKLKGFDICDCKIALDSGLSTGGFTDCLLQHG 150
QY 110 ASKYAIDVKGQQLHEKLYTNEQVINIEGVNLRTASKDLIPEEVDILTIDVSFISLTIL 169
Db 151 ASHVYGVGVYGVQVAEKIRTHERSVIERNLRYLSQ--LPEPVDLVTLDLSFISILVNM 208
QY 170 ESCRWLKASGIIIALIKPQFELYDPDKI-KKGWVKETSLQYEAVEKIIHFQCSBELGLIFI 228
Db 209 PAVIKVMTDSTLITLTKPFARRSQVGGGGIVRDLPHQEVLDRIISGVE-EFGFCNK 267
QY 229 GVPVSVIKGPKGNQEYL 245
Db 268 GWIESPIKGAEGNKEFL 284

RESULT 4
US-10-424-599-235806
; Sequence 235806, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 315806
; LENGTH: 312
; TYPE: PRT
```



```

; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 14023
; LENGTH: 271
; TYPE: PRT
; ORGANISM: Streptomyces avermitilis
; US-10-156-761-14023

Query Match 27.0%; Score 343.5; DB 14; Length 271;
Best Local Similarity 36.4%; Pred. No. 1.9e-29;
Matches 94; Conservative 48; Mismatches 87; Indels 29; Gaps 7;

QY 7 RADELVFLQGLAESRQAKRLIMAGKVTL-----TNNSTTIPRLKPKGKHYPLESICSL 61
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : | : :
Db 8 RLDAELVRRKLARSREHAGLIIAAGRVTGKTLATKSAQV-----ETAAAI 54
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : | : :

QY 62 IGVE-----RFVSRGAYKLLTALDF-----KIDVKSCICLDAGASTGGTDCLLQHGASKV 113
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : | : :
Db 55 LVTQDSDPDYVSRGKHLAGALAAFPQGLTVGRRALDAGASTGGTDFVLLRAGVAHV 114
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : | : :

QY 114 YAIIDVGKGLHEKLYNEQVINIEGVNLTASKDLP-EVDILTIDVFSISLTLLPSC 172
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : | : :
Db 115 VAUDVGYGQLAWSLQSDERVTVKDRNVNRELTLAIDGFPVDLVGDLSPFLGLVLPAL 174
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : | : :

QY 173 IRLKASGIIIALIKPQFELYDPKI-KKGVVKETSLOQFAVEKIIHFQSELGLIFIGVV 231
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : | : :
Db 175 VRCTAPCADLVMMVKPQPEVGKRLGGVVRSPFLRAEAVRGVAGRAW-ELGLGVKGYT 233
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : | : :

QY 232 PSVIKPGKNOEYLIYK 249
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : | : :
Db 234 ASLPGPGSNVEYFLWL 251
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : | : :

RESULT 8
US-10-335-977-7740
; Sequence 7740, Application US/10335977
; Publication No. US20040052799A1
; GENERAL INFORMATION:
; APPLICANT: DOUGLAS SMITH et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES
; RELATING TO HELICOBACTER PYLORI FOR
; DIAGNOSTICS AND THERAPEUTICS
; NUMBER OF SEQUENCES: 10031
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02109-1875
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD-ROM ISO9660
; COMPUTER: IBM PC Compatible
; OPERATING SYSTEM: Windows NT 4.0
; SOFTWARE: UNIX
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/335,977
; FILING DATE: 30-Dec-2002
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/993,002
; FILING DATE: 17-DEC-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Mandragouras, Amy E.
; REGISTRATION NUMBER: 36,207
; REFERENCE/DOCKET NUMBER: GTN-018
; TELECOMMUNICATION INFORMATION:

; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: (617)227-7400
; TELEFAX: (617)742-4214
; INFORMATION FOR SEQ ID NO: 7740:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 235 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: YES
; ORIGINAL SOURCE:
; ORGANISM: Helicobacter pylori
; FEATURES:
; NAME/KEY: misc feature
; LOCATION: (B) LOCATION 1...235
; SEQUENCE DESCRIPTION: SEQ ID NO: 7740:
US-10-335-977-7740

Query Match 24.2%; Score 307.5; DB 12; Length 235;
Best Local Similarity 35.2%; Pred. No. 1.6e-25;
Matches 87; Conservative 40; Mismatches 105; Indels 15; Gaps 5;

QY 6 VRADLVFLQGLAESRQAKRLIMAGKVTLTNNSTTIPRLKPKGKHYPLESICSLIGVE 65
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : | : :
Db 1 MRLDVALFNOHLANSREKAKALVKKQVLNKKVSKSFIVKGDQ-----IELIAPN 54
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : | : :

QY 66 RFVSRGAYKLLTALDFPKIDVKSCICLDAGASTGGTDCLLQHGASKVYVIDVGKQLHE 125
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : | : :
Db 55 LFSVRAGEKLGAFLEHDFIDFKKVVLDVGASKGGFSQALLKGAKKVLCDVGVKWLDE 114
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : | : :

QY 126 KLYTNEQVINIEGVNLRASKDLIPREVDILTIDVFSISLTLLPSCIRLWKASIIIAL 185
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : | : :
Db 115 SLKNDQRIEYCECDIRGFK--TPEKIDLCDFSFISLYCILEAI---LPLSGEFTL 168
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : | : :

QY 186 IKPQFEL--YDPKIKGVVVKETSLOQFAVEKIIHFQSELGLIFIGVPSVLKPKGNQE 243
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : | : :
Db 169 FKPDQFEVGRITKRNKKGVVMDKEATLNALNPFKNHLKTK-DQILTIQESLVKGNKNVE 227
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : | : :

QY 244 YLIYLYKK 250
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : | : :
Db 228 FFIHFGR 234
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : | : :

RESULT 9
US-10-335-977-7741
; Sequence 7741, Application US/10335977
; Publication No. US20040052799A1
; GENERAL INFORMATION:
; APPLICANT: DOUGLAS SMITH et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES
; RELATING TO HELICOBACTER PYLORI FOR
; DIAGNOSTICS AND THERAPEUTICS
; NUMBER OF SEQUENCES: 10031
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD
; STREET: 28 State Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02109-1875
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD-ROM ISO9660
; COMPUTER: IBM PC Compatible
; OPERATING SYSTEM: Windows NT 4.0
; SOFTWARE: UNIX
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/335,977
; FILING DATE: 30-Dec-2002
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/993,002
; FILING DATE: 17-DEC-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Mandragouras, Amy E.
; REGISTRATION NUMBER: 36,207
; REFERENCE/DOCKET NUMBER: 36,207

```

REFERENCE/DOCKET NUMBER: GTN-018
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617)227-7400
TELEFAX: (617)742-4214
INFORMATION FOR SEQ ID NO: 7741:
SEQUENCE CHARACTERISTICS:
LENGTH: 235 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: YES
ORIGINAL SOURCE:
ORGANISM: Helicobacter pylori
FEATURE:
NAME/KEY: misc_feature
LOCATION: (B) LOCATION 1...235
SEQUENCE DESCRIPTION: SEQ ID NO: 7741:
US-10-335-977-7741

Query Match 24.2%; Score 307.5; DB 12; Length 235;
Best Local Similarity 35.2%; Pred. No. 1.6e-25;
Matches 87; Conservative 40; Mismatches 105; Indels 15; Gaps 5;
QY 6 VRADELVFLQGLAESRQAKRLIMAGKVTLTNNSTTIPLRLKPKGKYPLESICSLIGVE 65
DB 1 MRLDYALFNQHLANSREKAKALVKKQVLYNKMVSPKSFIVKSGDQ-----IELIAPN 54
QY 66 RFVSRGAYKLLTALDPFKIDVKSCICLDAGASTGGTDCLLQHGASKYVAIDVKGQIHE 125
DB 55 LFVSRAGEKLGAFLEDHFIDFKVKVLDVGASKGFSQVALLKGAKKVLCDVVGKMQLDE 114
QY 126 KLYTNEQVINIEGVNLTASKDLIPEEVDILTIDVSPISLTILPSCIRWLKASGIIIAL 185
DB 115 SLKNDQRIEYECEDIRGFK---TPEKIDALCDVFSISLYCILEAL---LPLSGEFLTL 168
QY 186 IKPOFEL--YPDKIKKGWVKETSQYBAVEKIIHFCOSELGLIFIGVVPSPVKGPKGNQE 243
DB 169 FKPOFEVGRITKRNKKGVMVMDKEALNALENAFKNHLTK-DFOILTIOESLVKGNKNVE 227
QY 244 YLIYLKK 250
DB 228 FFIHFKR 234

RESULT 10

US-10-335-977-7742
Sequence 7742, Application US/10335977
Publication No. US20040052799A1
GENERAL INFORMATION:
APPLICANT: DOUGLAS SMITH et al
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES
RELATING TO HELICOBACTER PYLORI FOR
DIAGNOSTICS AND THERAPEUTICS
NUMBER OF SEQUENCES: 10031
CORRESPONDENCE ADDRESS:
ADDRESSEE: LAHIVE & COCKFIELD
STREET: 28 State Street
CITY: Boston
STATE: Massachusetts
COUNTRY: USA
ZIP: 02109-1875
COMPUTER READABLE FORM:
MEDIUM TYPE: CD-ROM ISO9660
COMPUTER: IBM PC Compatible
OPERATING SYSTEM: Windows NT 4.0
SOFTWARE: UNIX
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/335,977
FILING DATE: 30-Dec-2002
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/993,002
FILING DATE: 17-DEC-1997
ATTORNEY/AGENT INFORMATION:

NAME: Mandragouras, Amy E.
REGISTRATION NUMBER: 36,207
REFERENCE/DOCKET NUMBER: GTN-018
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617)227-7400
TELEFAX: (617)742-4214
INFORMATION FOR SEQ ID NO: 7742:
SEQUENCE CHARACTERISTICS:
LENGTH: 246 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: YES
ORIGINAL SOURCE:
ORGANISM: Helicobacter pylori
FEATURE:
NAME/KEY: misc_feature
LOCATION: (B) LOCATION 1...246
SEQUENCE DESCRIPTION: SEQ ID NO: 7742:
US-10-335-977-7742

Query Match 24.2%; Score 307.5; DB 12; Length 246;
Best Local Similarity 35.2%; Pred. No. 1.7e-25;
Matches 87; Conservative 40; Mismatches 105; Indels 15; Gaps 5;
QY 6 VRADELVFLQGLAESRQAKRLIMAGKVTLTNNSTTIPLRLKPKGKYPLESICSLIGVE 65
DB 12 MRLDYALFNQHLANSREKAKALVKKQVLYNKMVSPKSFIVKSGDQ-----IELIAPN 65
QY 66 RFVSRGAYKLLTALDPFKIDVKSCICLDAGASTGGTDCLLQHGASKYVAIDVKGQIHE 125
DB 66 LFVSRAGEKLGAFLEDHFIDFKVKVLDVGASKGFSQVALLKGAKKVLCDVVGKMQLDE 125
QY 126 KLYTNEQVINIEGVNLTASKDLIPEEVDILTIDVSPISLTILPSCIRWLKASGIIIAL 185
DB 126 SLKNDQRIEYECEDIRGFK---TPEKIDALCDVFSISLYCILEAL---LPLSGEFLTL 179
QY 186 IKPOFEL--YPDKIKKGWVKETSQYBAVEKIIHFCOSELGLIFIGVVPSPVKGPKGNQE 243
DB 180 FKPOFEVGRITKRNKKGVMVMDKEALNALENAFKNHLTK-DFOILTIOESLVKGNKNVE 238
QY 244 YLIYLKK 250
DB 239 FFIHFKR 245

RESULT 11

US-09-882-227-246
Sequence 246, Application US/09882227
Publication No. US20030158396A1
GENERAL INFORMATION:
APPLICANT: Kleanthous, Harold
APPLICANT: Al-Garawi, Amal
APPLICANT: Miller, Charles
APPLICANT: Tomb, Jean-Francois
APPLICANT: Oomen, Raymond P.
TITLE OF INVENTION: Identification of Polynucleotides
RELATING TO HELICOBACTER PYLORI
TITLE OF INVENTION: Encoding No. US20030158396A1
TITLE OF INVENTION: Genome
FILE REFERENCE: 06132/047002
CURRENT APPLICATION NUMBER: US/09/882,227
CURRENT FILING DATE: 2001-06-15
PRIOR APPLICATION NUMBER: US 08/902,615
PRIOR FILING DATE: 1997-07-29
NUMBER OF SEQ ID NOS: 638
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 246
LENGTH: 235
TYPE: PRT
ORGANISM: Helicobacter pylori
US-09-882-227-246
Query Match 22.9%; Score 290.5; DB 10; Length 235;

Best Local Similarity	33.6%;	Pred. No. 1.2e-23;	
Matches	83;	Conservative 41;	Mismatches 108; Indels 15; Gaps 5;
QY	6	VRDELVFLQGLAESRQAKRLIMAGKVTLLNNSTTIPRLLEKPGHKYPLESICSLIGVE	65
Db	1	MRLLDYALFSLHVNLSREKAKALVLKNQVLNVNQMVSFIVKENDK-----IELTAEK	54
QY	66	RFVSRGAYKLLTALDFFKIDVKSCICLDAGASTGGFTDCLIQHGASKVYAVIDVGKQIHE	125
Db	55	LFVSRAGEKLGAFLETHFVDFKGVVLVDVGASKGGFSQVALLKGAKRVLCDVVGKMLQDE	114
QY	126	KLYTNEOVINIEGVNLFTASKDLLIPREVDILTIOVSFISLTLLIIPSCIRWIKASGITIAL	185
Db	115	SLKQDKRIEYEECDIRGFK--TPETIDLALCDVSPISLYILEAI--LPLSDEFUTL	168
QY	186	IKQFQELYP--DKIKGVVAKETSQYEAIVEKI IHFCOSBELGLIIGVVPVSVIKGPKNGQE	243
Db	169	FKQFQFVGRGIRKNNKGVVDVKEAILNALENFKNHLTK-DFQILKTOESLVRKNGNVE	227
QY	244	YLIIYKK	250
Db	228	FFIIHFKR	234

```

RESULT 12
US-10-437-963-128703
; Sequence 128703, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 128703
; LENGTH: 161
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_31031C.1.pdp
US-10-437-963-128703

```

Query Match		18.9%;	Score 240;	DB 16;	Length 161;	
Best Local Similarity		37.3%;	Pred. No. 3.1e-18;			
Matches	53;	Conservative	38;	Mismatches	47;	Gaps 3;
QY	105	LICHGASKVYADVGVKGLHEKLYTNEQVINIEGVMLRTASKOLIPEEVDILIDVSFIS	164			
Ddb	4	LVPSNARGVIGDVGVGYVAEKIRHVERHSVIERTNLRYLTE--LPQLVDVLTDLSFIS	61			
		.: .: : .				
QY	165	LTLIPLSCIRWLKASGIITALLIKPQELYPDKT-KKGVVVKETSLQVEAEVKIIHFCSSEL	223			
Ddb	62	ILLVMFAVVKVMKTSTLTLLIKPQEARRSQVGCGIVRDLPHVKVELDRIISGVE-QF	120			
		.: .				
QY	224	GLIFIGVPDSVIKGPKNQEYL	245			
Ddb	121	GFCNKGWIESPIKGAEGNKEFL	142			
		.: .				

RESULT 13
US-10-425-114-56226
; Sequence 56226, Application US/10425114
; Publication NO. US2004003488A1
; GENERAL INFORMATION:

```

; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 56226
; LENGTH: 131
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: 700422554_FLI pep
US-10-425-114-56226

```

Query Match	15.5%	Score 198;	DB 12;	Length 131;
Best Local Similarity	39.7%;	Pred. No. 1.1e-13;		
Matches	46;	Conservative 28;	Mismatches 38;	Indels 4; Gaps 3;

QY	131	EQVINTEGNLRPSKXDLIPPEVDIITIDVSFSLTLLILPSCIRWLKASGIIITALIKPQF	190
	:	: :: :	: :: :
Db	1	ERVSVIERTNLRLSQ--LPPEVDLTLDLSFIILLVPAAVKMTDSTLTIKRPQF	58
	:	: :: :	: :: :
QY	191	ELYPDKI-KGVTVKTSLQVEAVEKLIHFCSQSELGLIFIGVPSVIKPGKNQBYL	245
	:	: :: :	: :: :
Db	59	EARRSVGGGIVRDPVLVHQEVLDRISGVE-EFGFCNKGIWESPIKAGSNGKEFL	113
	:	: :: :	: :: :

```

RESULT 14
US-10-424-599-183840
; Sequence 183840, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 183840
; LENGTH: 374
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_137021C.1.pep
US-10-424-599-183840

```

Query Match	8.0%;	Score 101.5;	DB 12;	Length 374;
Best Local Similarity	25.0%;	Pred. No. 0.031;		
Matches	44;	Conservative 26;	Mismatches 73;	Indels 33; Gaps 8;

QY	92	LDAGASTGCTDCLLOHGASKVYAIDVGKQGLHEKYTN----	EQVIN--EGVNLR	TASK	146
		: : : :	:	:	:
Db	66	LVDVGTGGILATWSAQAGARKVYAVETRMSEHARALVKANNIQQDVVEVIEG-----SMEE	121		
		: : :	:	:	:
QY	147	DLIPEVDILTIDVS--FISLTLLPSCSI-----RWLKASGIITIALIKPOPELYPD----	195		
		: : :	:	:	:
Db	122	ITLPEKVDDVIISEWGYFLLRSMFDSVINARDRWLKPTEGV-----MVPSHARMW	171		
		: : :	:	:	:
QY	196	--KIKKGVVKTSLQYEAWEKIHFHCQSHELIFIGVPSVIKPGKNQBYLYLK	249		
		: : :	:	:	:
Db	172	MAPIRGTIVDHKLGDYESMTMDWHFVEDT-KTYYGVDMLSTRPRFSSEEQRKKYLQ	226		
		: : :	:	:	:

RESULT 13
US-10-425-114-56226
; Sequence 56226, Application US/10425114
; Publication NO. US2004003488A1
; GENERAL INFORMATION:


```
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_50618C.1.pep
US-10-437-963-150375

Query Match      7.3%; Score 93; DB 16; Length 725;
Best Local Similarity 24.2%; Pred. No. 0.73;
Matches 64; Conservative 42; Mismatches 99; Indels 60; Gaps 14;

Qy 3 KHVRDELVFLOGLAESR-----EQAKRLIMAGKVTITNNSTTI-----42
Db 284 RHRAALDCCVQMLGLADRLADAAGAPDVEDVDARTWLSA---VLTDHVTCLDGLDDG 340
Qy 43 PLRLEKPHKYPERSICSL-IGVERFVSRYGAYKLLT-ALDPFK--IDVKSICICLDAGA-- 96
Db 341 PLRDSVGAHLEPLKSLASASIAVLSAAGRGRADVLAEAVDPFWLWLTARDTLLDAGAGA 400
Qy 97 -----STGGFT-----DCLQHGASKVYAIIDVGKGLHEKLYT--NEOVINIEG 138
Db 401 VQADVVAKDGSGKYTTIKEAVDAAPDGGKSR-YVIYVKGKYKENLEVGTKEKVLIVG 459
Qy 139 VNLTASKDLPIEVDILTIDVFSISLTL-----IIPSCIRWUKASGIIIALIKPOFELYPD 195
Db 460 DGM---DQTVITGSRNVVDGTTFNSATLGOKAVKASIA-LSGDGIIILQDLKVENTAGAE 515
Qy 196 KIKKGWVKETSLQYEAVEKIIHFQ 220
Db 516 K-----QQAVALRVSAADRAVINRCR 535

RESULT 20
US-10-282-122A-60355
; Sequence 60355, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
```

```
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 60355
; LENGTH: 444
; TYPE: PRT
; ORGANISM: Listeria monocytogenes
US-10-282-122A-60355

Query Match      7.3%; Score 92.5; DB 12; Length 444;
Best Local Similarity 22.0%; Pred. No. 0.4;
Matches 56; Conservative 34; Mismatches 87; Indels 77; Gaps 9;

Qy 16 GLAESREQAKRLIMAGKVTITNNSTTI-----LLEKPG-- 50
Db 162 GIEQLREIGLAFLVAPHQSIRVNVQTEIPTQLIKELNDQGITVTRNEFIEBALLVEKGSV 221
Qy 51 --HKYPLESTCSLIGVERFVSRYGAYKLLTALDFFKIDVKSCICLDAGASTGGTDCLLQ-- 107
Db 222 AETKAYDKGKCSIQDESSMLAAVALQLEDNL-----TVLDACAAPGKGTTHIAEK 271
Qy 108 -HGASKVYAIIDV--GKGLHEKLYTNEOVINIEGWL--RTASKDLIPEEVDILTIDVSF 162
Db 272 MHGTGMWHALDIEHKTKLIDEAARKRLQLNIRTAHLDAKTASTMFEFETFDRLVDA-- 329
Qy 163 ISLTLLPSCIRWUKASGIIIALIKPOFELYPDKIKGWVKETSLQYEAVEKIIHFQOSE 222
Db 330 -----PCSGFGVLRKRP-----DIKYAKTEKDIHKL-AE 357
Qy 223 LGLIFIGVPSVIK 236
Db 358 IQLAILDVDSQLVK 371

RESULT 21
US-10-072-012-554
; Sequence 554, Application US/10072012
; Publication No. US20040033493A1
; GENERAL INFORMATION:
; APPLICANT: Tchernev, Velizar
; APPLICANT: Spytek, Kimberly
; APPLICANT: Zertusen, Bryan
; APPLICANT: Patturajan, Meera
; APPLICANT: Shimkets, Richard
; APPLICANT: Li, Li
; APPLICANT: Gangolli, Esha
; APPLICANT: Padigaru, Muralidhara
; APPLICANT: Anderson, David W.
; APPLICANT: Rastelli, Luca
; APPLICANT: Miller, Charles E.
; APPLICANT: Gerlach, Valerie
; APPLICANT: Taupier Jr, Raymond J.
; APPLICANT: Gusev, Vladimir Y.
; APPLICANT: Colman, Steven D.
; APPLICANT: Wolenc, Adam R.
; APPLICANT: Pena, Carol E. A
; APPLICANT: Furtak, Katarzyna
; APPLICANT: Grosse, William M.
; APPLICANT: Alsbrook II, John P.
; APPLICANT: Lepley, Denise M.
; APPLICANT: Rieger, Daniel K.
; APPLICANT: Burgess, Catherine E.
; TITLE OF INVENTION: Proteins and Nucleic Acids Encoding Same
; FILE REFERENCE: 21402-258
; CURRENT APPLICATION NUMBER: US/10/072,012
; CURRENT FILING DATE: 2002-01-31
; PRIOR APPLICATION NUMBER: 60/265,102
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: 60/265,514
; PRIOR FILING DATE: 2001-01-31
; PRIOR APPLICATION NUMBER: 60/265,517
; PRIOR FILING DATE: 2001-01-31
; PRIOR APPLICATION NUMBER: 60/265,412
; PRIOR FILING DATE: 2001-01-31
```



```
; LENGTH: 562
; TYPE: PRT
; ORGANISM: Neisseria meningitidis
US-10-269-557-14

Query Match      6.9%; Score 87.5; DB 14; Length 562;
Best Local Similarity 23.7%; Pred. No. 2.1;
Matches 62; Conservative 35; Mismatches 86; Indels 79; Gaps 13;

QY 10 ELVFLQGLAERSREQAKRLIMAGKVTLTN-----NSTTIPRLRLEKPGHKYPLESICSLIGV 64
Db 176 DVNALQRLKEAEKAKIELSSGQTEINLPYITMDATGPKHLAMKITRAKFESL-----V 230
QY 65 ERFVSRGAYKLLTALDFFKIDVKSCICLDAGASTGGFTDCLLQHGASKVVAIDVGKQLH 124
Db 231 EDLITR-----SIEPCKIALK-----DAGLSTGDDIDVILVGGQSRMPKV----- 270
QY 125 EKLTYNEQVINIEGVNLRASKDLIPREV-----DILTIDVSFISLTL 167
Db 271 -----QEAVKAFGKEPR---KDVNPDEAVAGAAIQGEVLSSGRSDVLLDVTPLSLGI 322
QY 168 ILPSCIRWLKASGIIIIALIKPQELYPDKIKKGVVKETSQYEAIVEKIIHFQOSE----- 222
Db 323 -----ETMGGVMTKLIQKNTTI-PTKASQ-VFSTAEDNQSAV--TIHVLQGERERAS 370
QY 223 ----LGLIFIGVVPVSIKGPKG 240
Db 371 ANKSLGQFNLG---DIAPAPRG 389

RESULT 27
US-10-269-557-13
; Sequence 13, Application US/10269557
; Publication No. US20030099664A1
; GENERAL INFORMATION:
; APPLICANT: Winiowski, Jan
; TITLE OF INVENTION: HEAT SHOCK GENES AND PROTEINS FROM
; TITLE OF INVENTION: NEISSERIA MENINGITIDIS, CANDIDA GLABRATA AND ASPERGILLUS
; FILE REFERENCE: 870109.411
; CURRENT APPLICATION NUMBER: US/10/269,557
; CURRENT FILING DATE: 2002-10-11
; PRIOR APPLICATION NUMBER: US/09/207,388
; PRIOR FILING DATE: 1998-12-08
; NUMBER OF SEQ ID NOS: 102
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 13
; LENGTH: 642
; TYPE: PRT
; ORGANISM: Neisseria meningitidis
US-10-269-557-13

Query Match      6.9%; Score 87.5; DB 14; Length 642;
Best Local Similarity 23.7%; Pred. No. 2.5;
Matches 62; Conservative 35; Mismatches 86; Indels 79; Gaps 13;

QY 10 ELVFLQGLAERSREQAKRLIMAGKVTLTN-----NSTTIPRLRLEKPGHKYPLESICSLIGV 64
Db 256 DVNALQRLKEAEKAKIELSSGQTEINLPYITMDATGPKHLAMKITRAKFESL-----V 310
QY 65 ERFVSRGAYKLLTALDFFKIDVKSCICLDAGASTGGFTDCLLQHGASKVVAIDVGKQLH 124
Db 311 EDLITR-----SIEPCKIALK-----DAGLSTGDDIDVILVGGQSRMPKV----- 350
QY 125 EKLTYNEQVINIEGVNLRASKDLIPREV-----DILTIDVSFISLTL 167
Db 351 -----QEAVKAFGKEPR---KDVNPDEAVAGAAIQGEVLSSGRSDVLLDVTPLSLGI 402
QY 168 ILPSCIRWLKASGIIIIALIKPQELYPDKIKKGVVKETSQYEAIVEKIIHFQOSE----- 222
Db 403 -----ETMGGVMTKLIQKNTTI-PTKASQ-VFSTAEDNQSAV--TIHVLQGERERAS 450
QY 223 ----LGLIFIGVVPVSIKGPKG 240
Db 451 ANKSLGQFNLG---DIAPAPRG 469

RESULT 29
US-10-389-566-1292
; Sequence 1292, Application US/10389566
; Publication No. US20040025202A1
; GENERAL INFORMATION:
; APPLICANT: Monsanto Technology, LLC
; APPLICANT: Laurie, Cathy C
; TITLE OF INVENTION: Nucleic Acid Molecules Associated with Oil in Plants
; FILE REFERENCE: 38-77(52900)D
; CURRENT APPLICATION NUMBER: US/10/389,566
; CURRENT FILING DATE: 2003-03-31
; PRIOR APPLICATION NUMBER: US 60/365,301
; PRIOR FILING DATE: 2002-03-15
; PRIOR APPLICATION NUMBER: US 60/391,786
; PRIOR FILING DATE: 2002-06-25
; PRIOR APPLICATION NUMBER: US 60/392,018
; PRIOR FILING DATE: 2002-06-26
; NUMBER OF SEQ ID NOS: 2459
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 1292
; LENGTH: 642
; TYPE: PRT
```

```
; ORGANISM: Neisseria meningitidis
US-10-389-566-1292

Query Match
Best Local Similarity 6.9%; Score 87.5; DB 16; Length 642;
Matches 62; Conservative 35; Mismatches 86; Indels 79; Gaps 13;

QY 10 ELVFIQGLAESRQAKLIMAGKVTLTN-----NSTIPLRLKPGHKYPLESICSLIGV 64
DB 256 DVNALQRLKEAAEKIELSSGGQTEINLPYITMDATGPKHLAMKTRAKFESL-----V 310
QY 65 ERFVSRGAYKLLTALDFFKIDVKSCICLDAGASTGGFTDCLLOHGASKVYAIIDVKGQLH 124
DB 311 EDLITR-----SIEPCKIALK-----DAGLSTGDDIDVILVGGQSRMPKV----- 350
QY 125 EKLYTNEQVINIEGVNLRTRASKOLIPREV-----DILTIDVSFISLTL 167
DB 351 -----QEAUKAFPGKEPR---KDVNPDEAVAVGAIIQGEVLGGSRSDVLLDVTPLSLGI 402
QY 168 ILPSCIRWLKASGIIIALIKPQFELYPDKIKGVVKETSLSQYEAIVEKIIHFCQSE----- 222
DB 403 -----ETMGGVMTKLIQKNTTI-PTKASQ-VFSTAEDNQSAV--TIHVLQGERERAS 450
QY 223 ----LGLIFGVVPSVIKGPKG 240
DB 451 ANKSLGQFNLG---DIAPAPRG 469

RESULT 30
US-10-269-557-16
; Sequence 16, Application US/10269557
; Publication No. US20030095664A1
; GENERAL INFORMATION:
; APPLICANT: Wisniewski, Jan
; TITLE OF INVENTION: HEAT SHOCK GENES AND PROTEINS FROM
; TITLE OF INVENTION: NEISSERIA MENINGITIDIS, CANDIDA GLABRATA AND ASPERGILLUS
; TITLE OF INVENTION: FUMIGATUS
; FILE REFERENCE: 870109.411
; CURRENT APPLICATION NUMBER: US/10/269,557
; PRIOR FILING DATE: 2002-10-11
; PRIOR APPLICATION NUMBER: US/09/207,388
; PRIOR FILING DATE: 1998-12-08
; NUMBER OF SEQ ID NOS: 102
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 16
; LENGTH: 662
; TYPE: PRT
; ORGANISM: Neisseria meningitidis
US-10-269-557-16

Query Match
Best Local Similarity 6.9%; Score 87.5; DB 14; Length 662;
Matches 62; Conservative 35; Mismatches 86; Indels 79; Gaps 13;

QY 10 ELVFIQGLAESRQAKLIMAGKVTLTN-----NSTIPLRLKPGHKYPLESICSLIGV 64
DB 276 DVNALQRLKEAAEKIELSSGGQTEINLPYITMDATGPKHLAMKTRAKFESL-----V 330
QY 65 ERFVSRGAYKLLTALDFFKIDVKSCICLDAGASTGGFTDCLLOHGASKVYAIIDVKGQLH 124
DB 331 EDLITR-----SIEPCKIALK-----DAGLSTGDDIDVILVGGQSRMPKV----- 370
QY 125 EKLYTNEQVINIEGVNLRTRASKOLIPREV-----DILTIDVSFISLTL 167
DB 371 -----QEAUKAFPGKEPR---KDVNPDEAVAVGAIIQGEVLGGSRSDVLLDVTPLSLGI 422
QY 168 ILPSCIRWLKASGIIIALIKPQFELYPDKIKGVVKETSLSQYEAIVEKIIHFCQSE----- 222
DB 423 -----ETMGGVMTKLIQKNTTI-PTKASQ-VFSTAEDNQSAV--TIHVLQGERERAS 470
QY 223 ----LGLIFGVVPSVIKGPKG 240
DB 471 ANKSLGQFNLG---DIAPAPRG 489
```

```
RESULT 31
US-10-437-963-103572
; Sequence 103572, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 103572
; LENGTH: 959
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_100989C.1.pep
US-10-437-963-103572

Query Match
Best Local Similarity 6.9%; Score 87.5; DB 16; Length 959;
Matches 61; Conservative 25; Mismatches 73; Indels 103; Gaps 15;

QY 33 VTITNSTTPIRLKPGHKYPLESICSLIGVERFV--SRGAYKLLTALDFFKIDVKSCI 90
DB 591 VELTVGSKTIPTTF-----FVIDGKGSYLLGLRDM--IHANCCI 628
QY 91 CLDAGASTGGFTDCLLOHGASKVY-----AIDVGKGLHEKLYT 129
DB 629 -----PST--MNQLIQVQDKIEIVPADRSVNVASADLALWEMDGLDCLSGKMESPSY 681
QY 130 NEQVINTEGVNLRTRASKOLIP-----BEVDILTIDVS---FISLTLIP 170
DB 682 FEGV--VEGSNVYT--KDTVTDLDKQGGFMSADYLEIDIGFDRSRPTFISKNLSLE 737
QY 171 -----SCIRW--LKASGIIIAL-----IKPOFELY---PDKIKGVVKETS 206
DB 738 FRTKLIELKFRDCVDWEYEMSGLSRSIVEHRLPIKPGFRPHQPPRCCKADMLEPVK 797
QY 207 LQ-----YEAVEKIIHFQCSLQ 224
DB 798 AKIKELYDAGHKILSFMDSNAG 819

RESULT 32
US-10-282-122A-53447
; Sequence 53447, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
```


QY 201 --VVKETSLOYEAVEKEIIHF-----COSELGLIFGVVPSVIKGPKG 240
Db 210 HHLMEQSLLKYLKEGLAIFLAPVDILTSCSPLLKKWLQNNVTILAVISLPEAIFKQKS 269
QY 241 NOEYLIYLYKKR 251
Db 270 NIKSIFVLKKQ 280

RESULT 35

US-10-369-493-19647
; Sequence 19647, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 19647
; LENGTH: 644
; TYPE: PRT
; ORGANISM: Nitrosomonas europaea
US-10-369-493-19647

Query Match 6.8%; Score 87; DB 15; Length 644;
Best Local Similarity 23.3%; Pred. No. 2.9;
Matches 62; Conservative 36; Mismatches 78; Indels 90; Gaps 14;

QY 10 ELVFOGLAESREQAKRLIMAGKVTLTNNSTIP-LRLEKPGHKY-----PLESICS 60
Db 255 DMLALQRLKDAEAKAKIELSSQQTENV-----LPYITADAGPKHLAVKI TRAKLESU-- 308
QY 61 LIGVERFVSRGAYKILITALDFKIDVKSCICLDAGASTGGFTDCILLOHGASKVVAIDVGK 120
Db 309 ---VEELIERTAGPCRTALK-----DAGLSVSDINDVILVGGQTRMPKV----- 349
QY 121 GOLHEKLYTNQVNIENGVNLTASKDLIPEEV-----DILITDVSFI 163
Db 350 -----QEKVKEIFGKEPR--KQVNPDEAVAIGAIAQGGVLKGDVKDVLILLDVTPL 397
QY 164 SL---TLILPSCIRWLKASGIIIALIKPQFELYPDKIKKGVVYKETSLOYEAVEKEIIHFQ 220
Db 398 SLGIETL-----GGVWTKLIQKNTII-PTKAQQ-VFSTADDNQTA V--TIHVLQ 442
QY 221 SE-----IGHIFIGVWPSVIK 237
Db 443 GEREVASGNKSLGQFNLTIDIPSAPRG 468

RESULT 36

US-10-177-293-315
; Sequence 315, Application US/10177293
; Publication No. US20030124128A1
; GENERAL INFORMATION:
; APPLICANT: Lillie, James
; APPLICANT: Glatt, Karen
; APPLICANT: Zhao, Xumei
; APPLICANT: Gannavarpu, Manjula
; APPLICANT: Kamatkar, Shubhangi
; APPLICANT: Mertens, Maureen
; APPLICANT: Myer, Vic
; APPLICANT: Wang, Youzhen
; APPLICANT: Xu, Yongyao
; APPLICANT: Hoersch, Sebastian

; APPLICANT: Monahan, John
; APPLICANT: Meyers, Rachel E.
; APPLICANT: Baet Jr., Robert C.
; APPLICANT: Hortobagyi, Gabriel N.
; APPLICANT: Pusztai, Lajos
; APPLICANT: Meric, Funda
; APPLICANT: Sahin, Aysegul
; APPLICANT: Mills, Gordon B.
; TITLE OF INVENTION: COMPOSITIONS, KITS, AND METHODS FOR IDENTIFICATION, ASSESSMENT,
; TITLE OF INVENTION: PREVENTION, AND THERAPY OF BREAST CANCER
; FILE REFERENCE: MRI-038
; CURRENT APPLICATION NUMBER: US/10/177,293
; CURRENT FILING DATE: 2002-06-21
; PRIOR APPLICATION NUMBER: US 60/299,887
; PRIOR FILING DATE: 2001-06-21
; PRIOR APPLICATION NUMBER: US 60/301,572
; PRIOR FILING DATE: 2001-06-27
; PRIOR APPLICATION NUMBER: US 60/306,501
; PRIOR FILING DATE: 2001-07-18
; PRIOR APPLICATION NUMBER: US 60/325,002
; PRIOR FILING DATE: 2001-09-25
; PRIOR APPLICATION NUMBER: US 60/362,585
; PRIOR FILING DATE: 2002-03-05
; PRIOR APPLICATION NUMBER: US 60/xxx,xxx
; PRIOR FILING DATE: 2002-05-14
; NUMBER OF SEQ ID NOS: 506
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 315
; LENGTH: 1855
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-177-293-315

Query Match 6.8%; Score 86; DB 14; Length 1855;

Best Local Similarity 21.6%; Pred. No. 18;
Matches 61; Conservative 46; Mismatches 93; Indels 82; Gaps 15;

QY 2 AKHVRADLVFLOGLAESREQAKRLIMAGKVTLTN-----NSTTIP-LRLEKPGHKY 53
Db 302 AKENAHTRQACTILIGISEHQMGIFRILAGLHLGNVFTSRDADSCIP-----PKHE- 355
QY 54 PLESICSLIGVERFVSRGAY-----KLLTALDFFKIDVKSCICLDAGASTGGFTD 103
Db 356 PLCIFCDLGMVD-----YEEMCHWLCHRKLATATETY---IKPISKLOATNA-----RD 401
QY 104 CLLQHGASKVY--ALDYKGGOLHE--KLYTNEQVINIEGVNLTASKDLIPEEVDLITD 159
Db 402 ALAKHIVAKLFENWIVDNVNQALHSVAKQHSFIGVLDIYGF-----TFE 445
QY 160 VSFISLTILPSCIRWLKASGIIIALIKPQFELYPDKIK-----KGVVYKETSLOYEAVEK 214
Db 446 INSE-----QFCINYANEK-----LQQQFNMHVFKLEQEYMKEQIPWTLIDFYDNQP 494
QY 215 IHHFQCSLGLIFIGVWPSVIKGPKNQE-----YLIYLYKK 250
Db 495 CINLIESKIGI--LDLLDECKMPKGTDDTWAQKLYNTHLANK 534

RESULT 37

US-10-424-599-144586
; Sequence 144586, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684

```
; SEQ ID NO 144586
; LENGTH: 221
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_101574C.1.pep
US-10-424-599-144586

Query Match      6.7%  Score 85.5; DB 12; Length 221;
Best Local Similarity 21.1%; Pred. No. 0.86;
Matches 44; Conservative 40; Mismatches 78; Indels 47; Gaps 8;

QY 67 FVSRGAYKLTALDFFKIDVKSCICLDAGASTGGTDCILQ-----HGASKVVAIDVKGK 121
Db 19 YVARSAPKLLQIQNHQKIISPGSSILDLCAPGGWLVACQSLGPPFRGGSVLGVDTKKV 78
QY 122 QL-----HEKL-----YNEQVINIEGVNLRTASKDLI 149
Db 79 KVPFLHCDSRVQIISADVTTLPHQRLKALSPKKGFSVILSDMCPVSGGTTKDA---L 135
QY 150 PEEVDILTIVSFISLTILIPS-CIRWLKASG-IIIALIKPQFELYDPDKIKGVVKETS- 206
Db 136 SFELGMRALDLALGSRTHLEPSCDGVGVKGGHIVIKLLESEDAKEINQISKPLFKTSW 195
QY 207 LQYEAKEKIIHFCQSELGLPIGVVPSVI 235
Db 196 LRPKATRP-----SSREIYLICQGLKPDAM 220

RESULT 38
US-10-282-122A-45182
; Sequence 45182, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1

; SEQ ID NO 45182
; LENGTH: 337
; TYPE: PRT
; ORGANISM: Acinetobacter baumannii
US-10-282-122A-45182

Query Match      6.7%  Score 85.5; DB 12; Length 337;
Best Local Similarity 24.1%; Pred. No. 1.6;
Matches 62; Conservative 36; Mismatches 92; Indels 67; Gaps 14;

QY 6 VRADLVFLOGLAES--REQAKRLIMAGKVTLLNNSTTIP---LRLKPKGHKYPLES--- 57
Db 95 LKIDQSVFLVGBKGGVVERAAKQLQSFQKILKLD SARHCQLWHLKIKETKIKPLESWLK 154
QY 58 -----ICSLIGV--ERFVSRGAYKLTALDFFKIDVKSCICLDAGASTGGTDC 103
Db 155 TYTVQVNEQELTICALPGVFSQTHLDVGTAVLLPYLN-----QVKSRIADFGCG-AGIIS 209
QY 104 CLIQHGASK--VVAIDVGKQLHEKYTNEQVINIEGV---NLR-----TASKOLIPBEVD 154
Db 210 CYLAKANSSNIHALDIDAFALQ---STEMTFSRNGIGSDQLRLQPVGTGIAD-APTELD 264
QY 155 ILT-----IDVSFISLTILIPS CIRWLKASGIIIALIKPQFELYDPDKIKGVVKETS 206
Db 265 AIVSNPPFHQGHINTYDASEGLCQNAKHLKASG-ELMWIVANRFLNYPILIEK----- 316
QY 207 LQYEAKEKIIHFCQSEL 223
Db 317 -----HFGQCEI 323

RESULT 39
US-09-752-385-12
; Sequence 12, Application US/09752385
; Patent No. US20020012919A1
; GENERAL INFORMATION:
; APPLICANT: Anderson, Burt E.
; Regnery, Russell L
; TITLE OF INVENTION: Nucleic Acids of Rochalimaea Henselae
; and Methods and Compositions for Diagnosing
; Rochalimaea
; Henselae and Rochalimaea Quintana Infection
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NEEDLE & ROSENBERG, P.C.
; STREET: 127 Peachtree Street, N.E., Suite 1200
; CITY: Atlanta
; STATE: Georgia
; COUNTRY: USA
; ZIP: 30303
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/752,385
; FILING DATE: 29-Dec-2000
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/525,310
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Spratt, Gwendolyn D.
; REGISTRATION NUMBER: 36,016
; REFERENCE/DOCKET NUMBER: 1414.624
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (404) 688-0770
; TELEFAX: (404) 688-9880
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 160 amino acids
; TYPE: amino acid
```


GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 7, 2004, 16:52:00 ; Search time 22.5149 Seconds
(without alignments)
1072.357 Million cell updates/sec

Title: US-10-009-919A-1
Perfect score: 1271
Sequence: 1 MAKHKVRADELVFLQGLAES.....PSVIKPGKQGVLYLKKR 251

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues
Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR 78:*
1: pir1.*
2: pir2.*
3: pir3.*
4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	517.5	40.7	258	H70367	hemolysin - Aquifex
2	500.5	39.4	267	H97155	probable rRNA meth
3	495.5	39.0	267	A72238	hemolysin - thermo
4	439.5	34.6	271	E98007	conserved hypotet
5	432	34.0	274	S76154	hypothetical prote
6	431	33.9	274	AF1245	conserved hypotet
7	426.5	33.6	251	F97802	hemolysin [impor
8	425	33.4	274	AB1608	conserved hypotet
9	418.5	32.9	251	B71660	hemolysin (tlyA) R
10	418	32.9	269	G69967	hemolysin homolog
11	408.5	32.1	272	B83997	hemolysin-like pro
12	384.5	30.3	269	H87078	cytotoxin/hemolysi
13	383.5	30.2	240	A43863	cytotoxin/hemolysi
14	376.5	29.6	268	E70502	probable tlyA prot
15	372.5	29.3	250	AB2667	hemolysin [impor
16	372.5	29.3	250	H97448	hemolysin [impor
17	365	28.7	242	AG1919	hypothetical prote
18	364	28.6	268	S72968	cytotoxin/hemolysi
19	363	28.6	243	B86732	conserved hypotet
20	355.5	28.0	271	T36885	probable membrane
21	352.5	27.7	253	AC3439	hemolysin [impor
22	352	27.7	238	C90533	hemolysin a [impor
23	351	27.6	243	E87413	hemolysin A [impor
24	341.5	26.9	247	C82891	hemolysin UU436 [i
25	316	24.9	198	F95139	hemolysin A, proba
26	307.5	24.2	235	F71944	probable hemolysin
27	298.5	23.5	263	A70163	hemolysin (tlyA) h
28	290.5	22.9	235	F64655	hemolysin - Helico
29	270	21.2	253	C81406	probable hemeolysi

30	117	9.2	208	2	H69522	conserved hypotet
31	112.5	8.9	349	2	H64456	hypothetical prote
32	106.5	8.4	257	2	C86784	pseudouridine synt
33	93.5	7.4	642	2	B81917	probable chaperrone
34	92.5	7.3	444	2	AF1302	RNA-binding Sun pr
35	92.5	7.3	446	2	AF1674	RNA-binding Sun pr
36	91.5	7.2	348	2	T26447	hypothetical prote
37	89	7.0	245	2	B86738	hypothetical prote
38	89	7.0	328	2	A64471	hypothetical prote
39	89	7.0	475	2	F70144	carboxyl-terminal
40	88.5	7.0	627	2	A41609	dnak-type molecula
41	88	6.9	191	2	B69506	cell division prot
42	87.5	6.9	383	2	AB6182	hypothetical prote
43	87.5	6.9	642	2	H81185	dnak protein NMB05
44	87	6.8	1853	1	A46761	myosin heavy chain
45	86.5	6.8	207	2	F97014	SAM-dependent meth

ALIGNMENTS

RESULT 1

H70367

Hemolysin - Aquifex aeolicus

C:Species: Aquifex aeolicus

C:Date: 08-May-1998 #sequence_revision 08-May-1998 #text_change 24-Sep-1999

C:Accession: H70367

R:Deckert, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, D.E.; Ove
V.

Nature 392, 353-358, 1998

A:Title: The complete genome of the hyperthermophilic bacterium Aquifex aeolicus.

A:Reference number: A70300; MUID:98196666; PMID:9537320

A:Accession: H70367

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-258 <ADP>

A:Cross-references: GB:AE000707; NID:G2983342; PIDN:AA06935.1; PID:G2983347; GB:AE00065;

A:Experimental source: strain VF5

C:Genetics:

A:Gene: tly

C:Superfamily: hemolysin homolog yqxC

Query Match 40.7%; Score 517.5; DB 2; Length 258;
Best Local Similarity 43.6%; Pred. No. 1.3e-37;
Matches 105; Conservative 57; Mismatches 72; Indels 7; Gaps 3;

Qy 6 VRADELVFLQGLAESREQAQLIMAGKVTLTNNSTTIPRLLEKPGHKYPLESICSLIGVE 65

Db 1 MRLDKYLTDKGIVPSREKAQAVIMAGQV-LVNGKV-----VDKPGYRLKGNKKEVVEKELP 54

Qy 66 RFVSRGAYKLITALDFFKIDVKSCICLDAGASTGGFTDCLLQHGASKVYVIDVGRGQLHE 125

Db 55 KYVSRGGKLEWAKRFSLDLKDVKVLVDGSGTGFTDCLQHGAKKVVAVDVGQGM DY 114

Qy 126 KLYTNEQVINTIEGNLRTASKDLIPEEVDITIDVSFTLSLTILPSCRTWLKASIIIAL 185

Db 115 KLRQDPRVLYEETDARELSEHVPEKVDLTICDVSFISSTKVLNPFVKFKEDGLLVL 174

Qy 186 IKPQFELYPDKKKGVVKEVTSLOVEAVKLIHFCOSELGLIFIGVVPVSVIKPGKQGVYL 245

Db 175 VKPQFELCPKVKGVVREKHKRQALQKVNFLK-ENGFRLLGVIKSPKPGTKGNREFF 233

Qy 246 I 246

Db 234 V 234

RESULT 2

H97155

Probable rRNA methylase, YQXC B. subtilis ortholog [imported] - Clostridium acetobutylic

C:Species: Clostridium acetobutylicum

C:Date: 14-Sep-2001 #sequence_revision 14-Sep-2001 #text_change 30-Sep-2001

C:Accession: H97155

R;Nolling, J.; Breton, G.; Omeichenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Lee, J.; Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.
J. Bacteriol. 183, 4823-4838, 2001
A;Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium Cld
A;Reference number: A96900; MUID:21359325; PMID:21359325
A;Accession: H97155
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-267 <KUR>
A;Cross-references: GB:AB001437; PIDN:AAK80035.1; PID:g15025064; GSPDB:GN00168
A;Experimental source: Clostridium acetobutylicum ATCC824
C;Genetics:
A;Gene: CAC2076
C;Superfamily: hemolysin homolog yqxC

Query Match 39.4%; Score 500.5; DB 2; Length 267;
Best Local Similarity 44.7%; Pred. No. 4.1e-36;
Matches 113; Conservative 49; Mismatches 82; Indels 9; Gaps 4;

QY 1 MAKHKVRADELVFLQGLAESREOAKRLIMAGKVTLTNNSTIPLRLKPGHKYPLESICS 60
Db 1 MSENKERLDVLLVEKGFESREKARASIMAGEIYVDD-----LRIDKCGQKVKVSSKVE 54
QY 61 LIGVER-FVSRGAYKLLTALDFKIDVKSCICLDAGASTGGFTDCLLQHGASKVYVADVG 119
Db 55 FRGEKMPYVSRGGYKLERAKSFGLGKDKVCFDLAGSTGGFTDCMLQNGASKVFAIDVG 114
QY 120 KGQHEKLYTNEQVINIEGVNLTASKDLIPPEVDILTIDVSFISLTILPSCIRWLKAS 179
Db 115 YGQPAWKLRTPRVVCMERTNVRVTPEDIGEFNGFASIDVSFISLKKVVPVNLKDD 174
QY 180 GIITAIKLPQFELYPDKI-KKGVKETSLQYEAKEIHFQCSQELGLIFIGVWPSVTKGP 238
Db 175 GEIVALLKPOFEAGREXVGRGVREPEETHIEVINTIVDELK-EMKLSILGITYSPIKGP 233
QY 239 KGNQEYLIVLYKKR 251
Db 234 EGNIEYLIVYFSKK 246

RESULT 3
hemolysin - Thermotoga maritima (strain MSB8)
C;Species: Thermotoga maritima
C;Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 21-Jul-2000
C;Accession: A72238
R;Neilson, K.E.; Clayton, R.A.; Gill, S.R.; Gwinn, M.L.; Dodson, R.J.; Haft, D.H.; Hickey Garrett, M.M.; Stewart, A.M.; Cotton, M.D.; Pratt, M.S.; Phillips, C.A.; Richardson, D.; C.M.
Nature 399, 323-329, 1999
A;Title: Evidence for lateral gene transfer between Archaea and Bacteria from genome seq
A;Reference number: A72200; MUID:99287316; PMID:10360571
A;Accession: A72238
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-267 <ARN>
A;Cross-references: GB:AB001802; GB:AE000512; NID:g4982133; PIDN:AAD36643.1; PID:g498214
A;Experimental source: strain MSB8
C;Genetics:
A;Gene: TM1576
C;Superfamily: hemolysin homolog yqxC

Query Match 39.0%; Score 495.5; DB 2; Length 267;
Best Local Similarity 47.2%; Pred. No. 1.1e-35;
Matches 116; Conservative 37; Mismatches 86; Indels 7; Gaps 3;

QY 5 KVRADLVFLQGLAESREOAKRLIMAGKVTLTNNSTIPLRLKPGHKYPLESICSIGV 64
Db 4 KXLDQVILRGVLGVEREAKVLLAGKV-LVNGE-----RVTKASKLPEDANVELLEE 57
QY 65 ERFVSRGAYKLLTALDFKIDVKSCICLDAGASTGGFTDCLLQHGASKVYVADVGQGLH 124
Db 58 PKVSVRGYKLESAPESFKIDVSGKVACDLAGSTGGFTDCLLQHGAKVYVADVGQGLH 117

QY 125 EKLVTNEQVINIEGVNLTASKDLIPPEVDILTIDVSFISLTILPSCIRWLKASGIIIA 184
Db 118 WKLANDPRVVVMEKVNARYLNPDLGKVDVVTCDVSFISLKKIIPALSRILKNIGDAVL 177
QY 185 LIKQFELYPDKI-KKGVKETSLQYEAKEIHFQCSQELGLIFIGVWPSVTKGPKNQOEY 244
Db 178 LVKQFEPAPRKFRVKIGIVKDPGVHLEVEE-IRKSLIENGFWVKGCCFSKIKGTGEGNTEY 236
QY 245 LIYLKK 250
Db 237 FFWVK 242

RESULT 4
E98007
conserved hypothetical protein spr1086 [imported] - Streptococcus pneumoniae (strain R6
C;Species: Streptococcus pneumoniae
C;Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 02-Nov-2001
C;Accession: E98007
R;Hoskins, J.A.; Alborn Jr., W.; Arnold, J.; Blaszcak, L.; Burgett, S.; DeHoff, B.S.;
e, R.; LeBlanc, D.J.; Lee, L.N.; Lefkowitz, E.J.; Lu, J.; Matsushima, P.; McAhren, S.;
V, P.; Sun, P.M.; Winkler, M.E.
J. Bacteriol. 183, 5709-5717, 2001
A;Authors: Yang, Y.; Young-Bellido, M.; Zhao, G.; Zook, C.; Baltz, R.H.; Jaskunas, S.R.
A;Title: Genome of the Bacterium Streptococcus pneumoniae Strain R6.
A;Reference number: A97872; MUID:21429245; PMID:11544234
A;Accession: E98007
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-271 <KUR>
A;Cross-references: GB:AE007317; PIDN:AAK9989.1; PID:g15458709; GSPDB:GN00174
C;Genetics:
A;Gene: spr1086
C;Superfamily: hemolysin homolog yqxC

Query Match 34.6%; Score 439.5; DB 2; Length 271;
Best Local Similarity 42.5%; Pred. No. 8.8e-31;
Matches 107; Conservative 41; Mismatches 95; Indels 9; Gaps 5;

QY 1 MAKHKVRADELVFLQGLAESREOAKRLIMAGKVTLTNNSTIPLRLKPGHKYPLESICS 60
Db 1 MAKERV--DVLAAYKQGLFETREQAKRGVMAGLVAVVINGE---RFDKPGKEIPDDTELK 54
QY 61 LIGVE-RFVSRGAYKLLTALDFKIDVKSCICLDAGASTGGFTDCLLQHGASKVYVADVG 119
Db 55 LKGEKLYSVSRGGYKLEKALQVFDLSVDGATTIDIGASTGGFTDVMQLQNSAKLVFAVDVG 114
QY 120 KGQHEKLYTNEQVINIEGVNLTASKDLIPPEVDILTIDVSFISLTILPSCIRWLKAS 179
Db 115 TNQLAWKLRODPRVSGMEQNFYAEKTDPEQBPSPASIDVSFISLILPALHRVLADQ 174
QY 180 GIITAIKLPQFELYPDKI-KKGVKETSLQYEAKEIHFQCSQELGLIFIGVWPSVTKGP 238
Db 175 GQVVALLKPOFEAGREQIGKNGIIRDAKHQNVLESVTAMA-VEAGSVLDDDFSPIQGG 233
QY 239 KGNQEYLIVLYKK 250
Db 234 HGNIEFLAYLKK 245

RESULT 5
S76154
hypothetical protein slr0950 - Synechocystis sp. (strain PCC 6803)
C;Species: Synechocystis sp.
A;Variety: PCC 6803
C;Date: 25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change 20-Jun-2000
C;Accession: S76154
R;Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.;
o, K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda
DNA Res. 3, 109-136, 1996
A;Title: Sequence analysis of the genome of the unicellular cyanobacterium Synechocystis
s.

A;Reference number: S74322; MUID:97061201; PMID:8905231

A;Accession: S76154

A;Status: nucleic acid sequence not shown; translation not shown

A;Molecule type: DNA

A;Residues: 1-274 <KAN>

A;Cross-references: EMBL:D90914; GB:AB001339; NID:g1653477; PIDN:BAA18413.1; PID:g165350

A;Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996

C;Superfamily: hemolysin homolog yqxC

Query Match 34.0%; Score 432; DB 2; Length 274;

Best Local Similarity 42.9%; Pred. No. 4e-30;

Matches 111; Conservative 39; Mismatches 89; Indels 20; Gaps 6;

QY 1 MAK-HKVRDELVFLOGLAESREQAKELIWAGKVITLNNSTTIPLRLKPGHKVPLESIC 59

Db 1 MAKADQRLDALLIVAKGLCSRALAQRLIRAGEVKVNNQ-----LVDKFGTILVMDVAV 54

QY 60 SLIGVERFVSRGAYKLLTALDFPKIDVKSCICLDAGASTGGFTDCLLQHGASKVYAI DVG 119

Db 55 ELAQRPYVSRGGEKALAKALTEPAIDVTGICLDGGISTGGFTDCLLQRGATKVYGV DVG 114

QY 120 KGQHEKLYTNEQVINIEGVNLRITASDILPEV-----DILTIDVSFISLTLLPSC 172

Db 115 YGQVAMKLRQDDRVILRERANFRY---LTPQDLYGDPWPDLGWMDLFSISLTKVMAPL 170

QY 173 IRWLKASGIIIALTKPOFELYPDKI-KKGVVKETSLOVEAVEKIIHFQCSQELGLIFIGV 231

Db 171 WSLAPPREVLLVVKPOFEVGREKIGKGVVDRDAQAQAEALDQVWRSGQS-LGWQFNGLT 229

QY 232 PSVIKPGKNGQEVLIYLKK 250

Db 230 FSPITGPAGNVEYLLWSQ 248

RESULT 6

AF1245

conserved hypothetical protein lmol366 [imported] - Listeria monocytogenes (strain EGD-e

C;Species: Listeria monocytogenes

C;Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 14-Dec-2001

C;Accession: AF1245

R;Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloecker,

D.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi, H.

D.; Jones, L.M.; Karst, U.

Science 294, 849-852, 2001

A;Authors: Kretz, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.; Mat

ok, C.; Schluter, T.; Smoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehland,

A;Title: Comparative genomics of Listeria species.

A;Reference number: AB1077; MUID:21537279; PMID:11679669

A;Accession: AF1245

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-274 <GLA>

A;Cross-references: GB:NC 003210; PIDN:CAC99444.1; PID:g16410782; GSPDB:GN00177

A;Experimental source: strain EGD-e

C;Genetics:

C;Superfamily: hemolysin homolog yqxC

Query Match

Best Local Similarity

Matches 109; Conservative

QY 1 MAKHKVRDELVFLOGLAESREQAKELIWAGKVITLNNSTTIPLRLKPGHKVPLESICS 60

Db 1 MTIKERADILVEQGLFETREKAKRAIMAGITVYRKEE-----RVDPKPGKIPIDSELQ 54

QY 61 LIGVER-FVSRGAYKLLTALDFPKIDVKSCICLDAGASTGGFTDCLLQHGASKVYAI DVG 119

Db 55 VKGQMPYVSRGGLKELAKLVDFDEVKDKMLDLCAGSTGGFTDCLALQNGARHVSALDVG 114

QY 120 KGQHEKLYTNEQVINIEGVNLR-TASKDLIPPEVDILTIDVSFISLTLLPSCIRWLKA 178

Db 115 YNQLAWKLRNDRVTVMERTNFRHKVPADFAEGLADFATIDVSFISLKLPLVLRITVLVT 174

QY 179 SGIIIALIKPOFELYPDKI-KKGVVKETSLOVEAVEKIIHFQCSQELGLIFIGVPSVIK 237

Db 175 GGDVMTLIKPOFAGREQVKGKGIIRDPAVHESVVEHIVQFALDN-GYDLMGLDYSPI TG 233

QY 238 PKGNQBYLYLKK 249

Db 234 GEGNIEFIAHLK 245

RESULT 7

F97802

hemolysin [imported] - Rickettsia conorii (strain Malish 7)

C;Species: Rickettsia conorii

C;Date: 30-Sep-2001 #sequence_revision 30-Sep-2001 #text_change 22-Oct-2001

C;Accession: F97802

R;Ogata, H.; Audic, S.; Renesto-Audiffren, P.; Fournier, P.E.; Barbe, V.; Samson, D.; Ro

Science 293, 2093-2098, 2001

A;Title: Mechanisms of Evolution in Rickettsia conorii and Rickettsia prowazekii.

A;Reference number: A97700; MUID:21442074; PMID:11557893

A;Accession: F97802

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-251 <KUR>

A;Cross-references: GB:AE006914; PIDN:AAL03360.1; PID:g15619922; GSPDB:GN00173

C;Genetics:

C;Superfamily: hemolysin homolog yqxC

Query Match

Best Local Similarity

Matches 98; Conservative

QY 5 KVRADLVFLQGLAESREQAKRLIMAGKV-----TLTNNSTTIPLRLKPGHK 52

Db 3 KIRDELVLQKGFVDTITARSLLIQGVNKHQKLIQKPGIOWNINTEIKVKL--PQHN 60

QY 53 YPLESICSLIGVERFVSRGAYKLLTALDFPKIDVKSCICLDAGASTGGFTDCLLQHGASK 112

Db 61 Y-----VSRGALKUALDYFKIDPENLVACIDIGSSTGGFTVEMLERKAKL 106

QY 113 VYAI DVGKQGLHEKLYTNEQVINIEGVNLRITASDILPEVDILTIDVSFISLTLLPSC 172

Db 107 IFADVGVGELHPKLRDNPHIKVLEKTNARYLTDKQIITPKDLIVCDASFISLTLLPTV 166

QY 173 IRWLKASGIIIALTKPOFELYPDKIKK-GVVKETSLOVEAVEKIIHFQCSQELGLIFIGV 231

Db 167 INLVKDCMLIALIKPOFEVEKEVEGGVKNPDLHKQVCDKIKDMLKEHNEKIFGII 226

QY 232 PSVIKPGKNGQEVLIYLKKR 251

Db 227 ASPLGAKNGQEFILCGKRK 246

RESULT 8

AB1608

conserved hypothetical protein lin1403 [imported] - Listeria innocua (strain Clip11262)

C;Species: Listeria innocua

C;Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 14-Dec-2001

C;Accession: AB1608

R;Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloecker,

D.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi, H.,

Science 294, 849-852, 2001

A;Authors: Kretz, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.; Mat

ok, C.; Schluter, T.; Smoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehland,

A;Title: Comparative genomics of Listeria species.

A;Reference number: AB1077; MUID:21537279; PMID:11679669

A;Accession: AB1608

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-274 <GLA>

A;Cross-references: GB:AL592022; PIDN:CAC96634.1; PID:g16413876; GSPDB:GN00178

A;Experimental source: strain Clip11262
C;Genetics:
A;Gene: lin1403
C;Superfamily: hemolysin homolog yqx

Query Match 33.4%; Score 425; DB 2; Length 274;
Best Local Similarity 43.7%; Pred. No. 1.6e-29;
Matches 110; Conservative 33; Mismatches 99; Indels 10; Gaps 5;
QY 1 MAKHKVRADLVFLQGLAESRQAKRLIMAGKVTLTNNSTTIPRLRKPGHKYPLESICS 60
Db 1 MTIKKERADILLVEQGLFEFRERAKRAIMAGIVYRKEE-----RVDKPEKTPADELQ 54
QY 61 LIGVER-FVSRGAYKLLTALDFKIDVKSICILDAGASTGGFTDCLLQHGASKVYALDVG 119
Db 55 VKGQMPYVSRGGLKLEKALQVFNDFVDKMLDLIDIGASTGGFTDCLQNGARHSYALDVG 114
QY 120 KGQLEKLYTNEQVINTEGVNLR-TASKDLIPSEVDILTIDVFSITLTLIPSCIRWLKA 178
Db 115 YNLQAWKLNRDERTVVMERTNFRHVTTPADFTGLEAEFATIDVFSISLKLILPVLRTVLVT 174
QY 179 SGIITLALIKPQELYPDKI-KGVVKETSLSQYEAWEKIIHFQCOSELGLIFIGVVPVSIKG 237
Db 175 GGDVMTLLIKPQFAGREQVGKGIIRPAPVAVHEAVENIALFALDN-QYDLMLGLDFSPITG 233
QY 238 PKGNQBYLLIYK 249
Db 234 GEGNIEFIAHLK 245

RESULT 9
B71660
hemolysin (tlyA) RP555 - Rickettsia prowazekii
C;Species: Rickettsia prowazekii
C;Date: 21-Nov-1998 #sequence_revision 21-Nov-1998 #text_change 03-Nov-2000
C;Accession: B71660
R;Andersson, S.G.E.; Zomorodipour, A.; Andersson, J.O.; Sierhertz-Ponten, T.; Alsmark, U.
Nature 396, 133-140, 1998
A;Title: The genome sequence of Rickettsia prowazekii and the origin of mitochondria.
A;Reference number: A71630; MUID:99039499; PMID:9823893
A;Accession: B71660
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-251 <AND>
A;Cross-references: GB:AJ235272; GB:AJ235269; NID:g3861033; PIDN:CAA15004.1; PID:g386110
A;Experimental source: strain Madrid E
C;Genetics:
A;Gene: tlyA; RP555
C;Superfamily: hemolysin homolog yqx

Query Match 32.9%; Score 418.5; DB 2; Length 251;
Best Local Similarity 37.0%; Pred. No. 5.4e-29;
Matches 95; Conservative 49; Mismatches 88; Indels 25; Gaps 3;
QY 5 KVRADLVFLQGLAESRQAKRLIMAGKVTLTNNSTTIPRLRKPGHKYP 54
Db 3 KIRLDYLLQKGLVTDIARSLITQGVNHNKHEKLIKGVKNRHDPIKVKLPQHNY- 61
QY 55 LESICSLIGVERFVSRGAYKLLTALDFKIDVKSICILDAGASTGGFTDCLLQHGASKVY 114
Db 62 -----VSRGALKLITALDYFKIDPQNLVICIDIGSGTGGFTFVLFERKAEILF 108
QY 115 AIDVGQGLHEKLYTNEQVINTEGVNLR-TASKDLIPSEVDILTIDVFSITLTLIPSCIR 174
Db 109 AVDVGVGELHSLRFPQIKLEKTNARYITDKQITTKPDILVCDASFISLTILTPLEN 168
QY 175 WLKASGIITLALIKPQELYPDKI-KGVVKETSLSQYEAWEKIIHFQCOSELGLIFIGVVP 233
Db 169 LAKEDCILLALIKPQFVKNKNEVNGGIITNPLIHOKVCOKIKNWLEQEHFHQIFGIILAS 228
QY 234 VIKGPKGNQBYLLIYK 250
Db 229 PILGTGKNKEFLICGR 245

RESULT 10
G69967

hemolysin homolog yqx - Bacillus subtilis
N;Alternate names: hypothetical protein 65 (ahrC 5' region)
C;Species: Bacillus subtilis
C;Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 20-Jun-2000
C;Accession: G69967; PS0053
R;Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Berte
C.; Bron, S.; Brouillet, S.; Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Ch
A.; Ehrlich, S.D.; Emmerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E.
Nature 390, 249-256, 1997
A;Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Galle
iech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.
Koeter, P.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois
A;Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maue
Y, M.; Ogawa, K.; Ogiwara, B.; Oudega, B.; Park, S.H.; Parro, T.; Pohl, T.M.; Portetell
Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadaie, Y.; Sato, T.; Scanlon
A;Authors: Schleich, S.; Schroeter, R.; Scoffone, P.; Sekiguchi, J.; Sekowska, A.; Seron
teuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiyama
T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, H.
A;Authors: Yoshikawa, H.F.; Zumstein, E.; Yoshikawa, H.; Danchin, A.
A;Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis.
A;Reference number: A69580; MUID:98044033; PMID:9384377
A;Accession: G69967
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-269 <KUN>
A;Cross-references: GB:Z99116; GB:AL009126; NID:g2634723; PIDN:CAB14357.1; PID:g2634860
A;Experimental source: strain 168
R;North, A.K.; Smith, M.C.M.; Baumberg, S.
Gene 80, 29-38, 1989
A;Title: Nucleotide sequence of a Bacillus subtilis arginine regulatory gene and homolog
A;Reference number: A91616; MUID:90006783; PMID:2507400
A;Accession: PS0053
A;Molecule type: DNA
A;Residues: 178-179; 'LIAQ', 211-269 <NOR>
A;Cross-references: GB:M27869; NID:g142448; PIDN:AAA22207.1; PID:g142449
A;Experimental source: strain EMG50
C;Genetics:
A;Gene: yqx
C;Superfamily: hemolysin homolog yqx

Query Match 32.9%; Score 418; DB 2; Length 269;
Best Local Similarity 43.4%; Pred. No. 6.5e-29;
Matches 108; Conservative 38; Mismatches 93; Indels 10; Gaps 5;
QY 1 MAKHKVRADLVFLQGLAESRQAKRLIMAGKVTLTNNSTTIPRLRKPGHKYPLESICS 60
Db 1 MTSKKERLDVLLAVERGLAETREKAKRAIMAGIVYSNEN-----RLDKPEKIDRDLPLT 54
QY 61 LIGVE-FVSRGAYKLLTALDFKIDVKSICILDAGASTGGFTDCLLQHGASKVYALDVG 119
Db 55 VKGNPFRVYSRGGLKLEKALKEFPVSVKDKIMIDIGSGTGGFTDCLQNGAKOSYALDVG 114
QY 120 KGQLEKLYTNEQVINTEGVNLR-TASKDLIPSEVDILTIDVFSITLTLIPSCIRWLKA 178
Db 115 YNLQAWKLQDQRRVVMERTNFRYATPADFTKGMPFATIDVFSISLRLILPVLRTLLVP 174
QY 179 SGIITLALIKPQELYPDKI-KGVVKETSLSQYEAWEKIIHFQCOSELGLIFIGVVPVSIKG 237
Db 175 GSDCMALVKPQFAGRESVGKGIIRDPKVDVLRKMLSPFSAAE-GYICKLSLFSFPIG 233
QY 238 PKGNQBYLLI 246
Db 234 GDNIEFLL 242

RESULT 11
B83997

hemolysin-like protein BH2778 [imported] - Bacillus halodurans (strain C-125)
C;Species: Bacillus halodurans

C;Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 15-Jun-2001
C;Accession: B83997
C;Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.; Hirai
Nucleic Acids Res. 28, 4317-4331, 2000
A;Title: Complete genome sequence of the alkaliphilic bacterium *Bacillus halodurans* and
A;Reference number: A83650; MUID:20512582; PMID:11058132
A;Accession: B83997
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-272 <STO>
A;Cross-references: GB:AP001516; GB:BA000004; NID:g10175192; PIDN:BA06497.1; GSPDB:GN00
A;Experimental source: strain C-125
C;Genetics:
A;Gene: BH2778
C;Superfamily: hemolysin homolog yqxc

Query Match 32.1%; Score 408.5; DB 2; Length 272;
Best Local Similarity 41.7%; Pred. No. 4.5e-28;
Matches 106; Conservative 48; Mismatches 89; Indels 11; Gaps 6;

QY 1 MAKHKVRADELVFLQGLAESRQAKRLIMAGKVTLTNNSTIPLRLKPGHKYPLESICS 60
Db 1 MSK-KERVDVLLVERGLMETREKAKRISIMAGLVFSGHE-----RVDKPGKLVDRDTPLS 53
QY 61 LIG-VERFVSRGAYKLLTALDFFKIDVKSCICLDAGASTGGTFTDCLLQHGASKVYVADVG 119
Db 54 VKEVLVPSRGGHLEKATRAFDLHITDRVLDIGASTGGTFTDCAQNGATFYVADVG 113
QY 120 KGQLEKLYTNEQVINIEGVNLTASKDLIPEEV-DILTTIDVSFISLTILPSCIRWLKA 178
Db 114 YNQLAWKLQDERVVMVMTNFRYLKPEVLRLGPNMATIDVSFISLKLILPVLKTMLE 173
QY 179 SGIITIALIKPQFELYPKI-KKGVVKETSLOYEAVEKLIHFCQSELGLIFIGVPSVIKG 237
Db 174 NSDVALVKQFQFAGREVEVGKGIIVRDKSVHQKVLSTIVEFALKE-GYAVGGLDFSPITG 232
QY 238 PKGNQBYLYLKKR 251
Db 233 GEGNIEFLHLWLR 246

RESULT 12
H87078
Cytotoxin/hemolysin [imported] - Mycobacterium leprae
C;Species: Mycobacterium leprae
C;Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 10-May-2001
C;Accession: H87078
R.; Cole, S.T.; Eiglmeier, K.; Parkhill, J.; James, K.D.; Thomson, N.R.; Wheeler, P.R.; Ho
R.; Davies, R.M.; Devlin, K.; Duthoy, S.; Feltwell, T.; Fraser, A.; Hamlin, N.; Holroyd,
eam, M.A.; Rutherford, K.M.
Nature 409, 1007-1011, 2001
A;Authors: Rutter, S.; Seeger, K.; Simon, S.; Simmonds, M.; Skelton, J.; Squares, R.; S
A;Title: Massive gene decay in the leprosy bacillus.
A;Reference number: A86909; MUID:21128732; PMID:11234002
A;Accession: H87078
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-269 <STO>
A;Cross-references: GB:AL450380; NID:g13093256; PIDN:CAC31739.1; GSPDB:GN00147
C;Genetics:
A;Gene: tlva
C;Superfamily: hemolysin homolog yqxc

Query Match 30.3%; Score 384.5; DB 2; Length 269;
Best Local Similarity 38.9%; Pred. No. 5.4e-26;
Matches 98; Conservative 49; Mismatches 94; Indels 11; Gaps 5;

QY 1 MAKHKVRADELVFLQGLAESRQAKRLIMAGKVTLTNNSTIPLRLKPGHKYPLESICS 60
Db 1 MAR-RVRVDVLRVRLGLARSQQAALKLSAGKVSIDG-----LPVAKPSTAVATITVLT 53
QY 61 LI--GVRFVSRGAYKLLTALDFFKIDVKSCICLDAGASTGGTFTDCLLQHGASKVYVADVG 118

Db 54 VADGERSWVRGAHKLMGALDTFGIPVAGRCCLDAGASTGGTEVLLDRGAEEVAVDV 113
QY 119 KGQLEKLYTNEQVINIEGVNLTASKDLIPEEV-DILTTIDVSFISLTILPSCIRWLKA 178
Db 114 GYQLAWSVRCDPRVIVERTNVHDLSPELIGPVDLVVADLSFISLATVLSALAGCALP 173
QY 179 SGIITIALIKPQFELYPKI-KKGVVKETSLOYEAVEKLIHFCQSELGLIFIGVPSVIKG 237
Db 174 SADIVPMKPKQFVEGKQVPGGVRDLRLRADSVLAVARRA-TELGWRTMDVTASSLP 232
QY 238 PKGNQBYLYLKK 249
Db 233 LSGNVEYFLMLR 244

RESULT 13
A43863
cytotoxin/hemolysin ORF2 tly - *Treponema hyodysenteriae*
C;Species: *Treponema hyodysenteriae*
C;Date: 04-Sep-1998 #sequence_revision 04-Sep-1998 #text_change 26-Aug-1999
C;Accession: A43863
R.; Muir, S.; Koopman, M.B.; Libby, S.J.; Joens, L.A.; Heffron, F.; Kusters, J.G.
Infect. Immun. 60, 529-535, 1992
A;Title: Cloning and expression of a *Serpula* (*Treponema*) hyodysenteriae hemolysin gene.
A;Reference number: A43863; MUID:92112315; PMID:1730486
A;Accession: A43863
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-240 <MUI>
A;Experimental source: strain B204
A;Note: sequence extracted from NCBI backbone (NCBIN:77501, NCBI:77504)
C;Superfamily: hemolysin homolog yqxc

Query Match 30.2%; Score 383.5; DB 2; Length 240;
Best Local Similarity 38.6%; Pred. No. 5.8e-26;
Matches 96; Conservative 41; Mismatches 93; Indels 19; Gaps 5;

QY 6 VRADELVFLQGLAESRQAKRLIMAGKVTLTNNSTIPLRLKPGHKYPLESICSLIGVE 65
Db 1 MELDEYVHSEGYTESRSKAQDIIILAGCVFNG-----VKVTSKAHKIKDTDNIEVVQNI 54
QY 66 RFVSRGAYKLLTALDFFKIDVKSCICLDAGASTGGTFTDCLLQHGASKVYVADVGQLHE 125
Db 55 KYVSRAGEKLEKAFVEFGISVENKICLDIGASTGGTFTDCLLQHGAKKYVYALDVGHNLVY 114
QY 126 KLYTNEQVINIEGVNLTASKDLIPEEV-DILTTIDVSFISLTILPSCIRWLKASGIIIA 184
Db 115 KLRNDNRVWSIEDFNADINKEMENDEIPSVIVSDVSFISITKAPIIFKELNNLEFWVT 174
QY 185 LIKQFELYPPDKIKKG-----VVKETSLOYEAVEKLIHFCQSELGLIFIGVPSVIKGP 238
Db 175 LIKQFPAERGDVSKGGIIRDDILREKILN-NAISKII-----DCGPFKVNRTISPIKGA 228
QY 239 KGNQBYLYIY 247
Db 229 KGNIEYLAH 237

RESULT 14
E70502
probable tlyA protein - *Mycobacterium tuberculosis* (strain H37RV)
C;Species: *Mycobacterium tuberculosis*
C;Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 20-Jun-2000
C;Accession: E70502
R.; Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.
R.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.;
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A;Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrrell, B.G.
A;Title: Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome
A;Reference number: A70500; MUID:98295987; PMID:9634230
A;Accession: E70502
A;Status: preliminary; nucleic acid sequence not shown; translation not shown

A;Experimental source: strain PCC 7120
C;Genetics:
A;Gene: all0906
C;Superfamily: hemolysin homolog yqxC

Query Match 28.7%; Score 365; DB 2; Length 242;
Best Local Similarity 43.6%; Pred. No. 2.4e-24;
Matches 92; Conservative 30; Mismatches 77; Indels 12; Gaps 5;

46	QY	LEPEGHKYPLESTCSLIGVERFVUSRGAYKLLTALDFFKIDVKYSCICLDAGASTGCGFTDCL	105
8	Db	IDRPGTEVDVAAQVKIKERSRFRVSRGGEKAKGLELFAISVNGRVCLDGGISTGCGFTDCL	67
106	QY	LQHGASKVYALDVGKGQJLHEKLYNEOVINTEGVNLRITAS-KDL-----IPEEVDILTI	158
68	Db	LQHGAKLVYIGDVGYGQTDGRLNRPQVILRERNLRQLQPKDLYGDCDAIP---DLAVV	124
159	QY	DVSPFISLTILPSCIRWLKASGIIIALIKPOFELYDPDKI-KKGVVKETSLOYEAVEKIIH	217
125	Db	DVSPFISLTILPALWQLTQPPREAVLLVKKQFEVKGSRVKGKGVVRDPHDQADAIQVWQ	184
218	QY	FCQSEGLIIFIGVVPVSVIKGPKGNQEVLIYL	248
185	Db	AAB-KLGWKYKGLTWSPIITGPAGNIEYLLWL	214

RESULT 18
S72968
cytotoxin/hemolysin homolog u0247a - Mycobacterium leprae
C;Species: Mycobacterium leprae
C;Date: 19-Mar-1997 #sequence revision 25-Apr-1997 #text change 23-Mar-2001

C/Accession: S72589
R/Smith, D.R.; Robison, K.
Submitted to the EMBL Data Library, November 1993
A/Description: Mycobacterium leprae cosmid L247.
A/Reference number: S72589

A;Accession: J02768
A;Status: Preliminary
A;Molecule type: DNA
A;Residues: 1-268 <SMI>
A;CROSS-references: EMBL:U00021; NID:G467141; PID:AAA50906.1; PID:G467142
C;Genetics:
C;Start codon: GTG
C;Superfamily: hemolysin homolog vxqC

Query Match. 28.6%; Score 364; DB 2; Length 268;
Best Local Similarity 38.2%; Pred. NO. 3.3e-24;
Matches 96; Conservative 48; Mismatches 97; Indels 10; Gaps 5

Qy		1	MAKHVRADELVFUOGLAESREQAKRLIMAGKVLTNNSTIPLRLEFGHKYPLESTCS	60
	:	:	:	:
	:	:	:	:
Db		1	MAR-RVRVDVELVRGSLRSQQAAKLSAGKVSIDG-----LPAPVPSTAVALTITVLIT	53
	:	:	:	:
	:	:	:	:
Qy		61	LI-GVERFVSRGAYKLTLTALDFFKFIDVKSCSICLDAGASTGGTFCDLLOHGASKKYVAIDVG	119
	:	:	:	:
	:	:	:	:
Db		54	VADDEGSWVCAGHLMGALDTFGIPVAGRCCLDAGASTGGTFEVLLDORGAEEVVAVDVG	113
	:	:	:	:
	:	:	:	:

Qy	120	KGQJHEKYTNEQVINTEGVNLRATASKDLIPEEVILFTIDVSFSLTILPSCIRWLKAS	179
Db	114	YGQJAWSVRCDPRIVIVERTNVHDLSPELIGPVDLVVADLSFSLATVLSALACALPS	173
Qy	180	GIITALKPQPELYPDKI-KGVVKTSLQYEAVEKIIHFCQSEGLGIFIGVWPSVIKGP	238
Db	174	ADIVPMVKPQPEVGKQGVGGVWRDLRLRADSVLAVARRA-TELGWRTMDVTASSLPCL	232
Qy	239	KCQOEYLIYLYK	249
Db	233	SGNVEYFLWLR	243

RESULT 19
B96732
conserved hypothetical protein viiB [imported] - Lactococcus lactis subsp. lactis (strain

C;Species: *Lactococcus lactis* subsp. *lactis*
C;Date: 23-Mar-2001 #sequence_revision 23-Mar-2001 #text_change 03-Aug-2001
C;Accession: B86732
R;Bolotin, A.; Winkler, P.; Manger, S.; Jaillon, O.; Malarne, K.; Weissenbach, J.; Ehrlich
Genome Res. 11, 731-753, 2001
A;Title: The complete genome sequence of the lactic acid bacterium *Lactococcus lactis* ssj
A;Reference number: A86625; MUID:21235186; PMID:11337471
A;Accession: B86732
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-243 <STC>
A;Cross-references: GB:AE005176; PID:gl1273786; PIDN:AAK04956.1; GSPDB:GN00146
A;Experimental source: strain Il1403
C;Genetics:
A;Gene: YiiB
A;Superfamily: hemolysin homolog vxvC

Query Match 28.6%; Score 363; DB 2; Length 243;
Best Local Similarity 40.4%; Pred. NO. 3.6e-24;
Matches 91; Conservative 40; Mismatches 86; Indels 8; Gaps 5;

Qy	29	MAGKVTLTNNS	TTIPLRELPKCHKYPLESICSLGVE-RFVSRGAYKLLTALDPFKIDVK	87
Db	1	MAGLVDSKSGE	-----RYDPGQKIDGDTTEURLKGEKLYVSRGSLKLEKAKSFHLEIN	56
Qy	88	SCICLDAGASTGGFTDCLLGHGASKVAIDVGKQLHEKLYTNEOVINIEGWNLRATA-SK	146	
Db	57	GKTCCLDGGSTGGFTDVMQLGAKLVVALDVTGTLQAKIRSDERVVVVMEQENFNAVULA	116	
Qy	147	DLIPEEVDILTIDVSFSLTLILPESCIRWLKASGIIIALIKPQPELYPKDI-KKGWVKET	205	
Db	117	DFAQGREFSFTSIDVSFISLDLILPLVEILEKNGEVAALIKPQPEAGREQVGKNGIIRDP	176	
Qy	206	SLQVEAVEKIIHFQCSBELGLIFIGWPSVIKPGKGNQYLYLVKK	250	
Db	177	KVHOMTTEKVLK-TATQLGFSVKGTGTFSPIKGGAGNVFELVHLKK	220	

RESULT 20
T36885

probable membrane protein - Streptomyces coelicolor
C/Species: Streptomyces coelicolor
C/Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 :
C/Accession: T36885
R/Murphy, L.; Harris, D.; James, K.D.; Parkhill, J.
submitted to the EMBL Data Library, August 1999
A/Reference number: Z21617

A;Accession: J00003
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 1-271 <MUR>
A;Cross-references: EMBL:AL10948; PIDN:CAB52846.1; GSPDB:GN00070; SCOEDB:SC151.22C
A;Experimental source: strain A3(2)
C;Genetics:
A;Gene: SCOEDB:SC151.22C
C;Superfamily: hemolysin homolog yqxC

Query Match	28.0%;	Score 355.5;	DB 2;	Length 271;
Best Local Similarity	37.5%;	Pred. No. 1.9e-23;		
Matches 95: Conservative	45;	Mismatches 94;	Indels 19;	Gaps 6;

Qy	7	RADELVFLOGLAESREOAKRLIMAGKVTLTNNSTTIPLRLEKPGHKYPLESTCSLI----	62
Db	8	RLDAELVRRKLARSREHASQLIAGRVTGVKTKPAT-----QVETAAAIWVTAD	59
Qy	63	-GVERFVSRGAYKLLTALDFF---KIDVSCICLDAGASTGGFTDCLLQHGASKYVADV	118
Db	60	DNDPDVYSRGGHKLAGALAAFPVHGLVVEGRRALDAGASTGGFTDVLLEAGAAHVAVDV	119
Qy	119	KGQQLHEKLYTNEQVINIEGVNLRATASKDLIP-BEVDILTIDVSPISLTLLIIPSCIRWLK	177
Db	120	GYGQLAWSLRDERVTWKDTRINVELTEPAIDGEPVDLVGGLSPILVALVPAIVLRCTR	179

```
QY 178 ASGIITAIKPOELYPDKI-KKGVVKETSLQYEAIVEKIHFQCOSELGLIFIGVVPVSIK 236
Db 180 PGADLVMMVVPQPEVGGKRLGSGVRSQAQLRAEAVRGVARKAW-ELGLGVKGVGTASPLP 238
QY 237 GPKGNQOEYLIYIK 249
Db 239 GPSGNVEYFLWLR 251

RESULT 21
AC3439
hemolysin [imported] - Brucella melitensis (strain 16M)
C;Species: Brucella melitensis
C;Date: 01-Feb-2002 #sequence_revision 01-Feb-2002 #text_change 06-Jan-2003
C;Accession: AC3439
R;DelVecchio, V.G.; Kapatal, V.; Redkar, R.J.; Patra, G.; Mijer, C.; Los, T.; Ivanova,
.; Mazur, M.; Goltzman, E.; Selkov, E.; Elzer, P.H.; Hagius, S.; O'Callaghan, D.; Letes
Proc. Natl. Acad. Sci. U.S.A. 99, 443-448, 2002
A;Title: The genome sequence of the facultative intracellular pathogen Brucella melitensis
A;Reference number: AD3252; PMID:11756688
A;Accession: AC3439
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-253 <KUR>
A;Cross-references: GB:AE008917; PIDN:AAU52678.1; PID:G17983504; GSPDB:GN00190
A;Experimental source: strain 16M
C;Genetics:
A;Gene: BME11497
A;Map position: 1
C;Superfamily: hemolysin homolog yqxc

Query Match 27.7%; Score 352.5; DB 2; Length 253;
Best Local Similarity 36.4%; Pred. No. 3.1e-23;
Matches 91; Conservative 43; Mismatches 107; Indels 9; Gaps 4;

QY 4 HKVRADELVLQGLABREQAARLIMAGKVTLTNNSTTIPRLKPKGHKYPLESICSLIG 63
Db 8 NRRLDQLIVERGFEATRSARDAIORGTGVKVGPRPT-----KPGQWVVRDAALAVDD 61
QY 64 -VERFVSRAAYKLLTALDFFKIDVKSCICLDAGASTGGFTDCLLQHGASKVYVIDVGKQ 122
Db 62 PASAYVSRAALKLVLAALDHLNKKVGTALDAGSTGGFTQVLLERGAHVIAIDVGHQD 121
QY 123 LHKLYTNEOVINIEGNLNTAS-KDLIPREVILITDVFSITLTLPSCIRMLKASGI 181
Db 122 LHESLRHDPRTVSKGVNARALELAHLDTRAVDCIVSDVFSISRLALPPALAEKAI 181
QY 182 IIALIKPOELYPDKI-KKGVVKETSLQYEAIVEKIHFQCOSELGLIFIGVVPVSIKPKG 240
Db 182 CALLVLPFLFAGREAIKGGILKDPAYGERMAQELKSWLETQPGWRALGLCPSPIEGGD 241
QY 241 NQYLIYIKK 250
Db 242 NREYLLAGKK 251

RESULT 22
C90533
hemolysin a [imported] - Mycoplasma pulmonis (strain UAB CTIP)
C;Species: Mycoplasma pulmonis
C;Date: 24-May-2001 #sequence_revision 24-May-2001 #text_change 03-Aug-2001
C;Accession: C90533
R;Chambaud, I.; Heilig, R.; Ferris, S.; Barbe, V.; Samson, D.; Galisson, F.; Moszer, I.;
Nucleic Acids Res. 29, 2145-2153, 2001
A;Title: The complete genome sequence of the murine respiratory pathogen Mycoplasma pulm
A;Reference number: A99512; MUID:21267165; PMID:11353084
A;Accession: C90533
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-238 <KUR>
A;Cross-references: GB:AL445566; PID:G14089584; PIDN:CAC13344.1; GSPDB:GN00153
A;Experimental source: strain UAB CTIP
C;Genetics:
```

```
A;Gene: MYPU 1710
A;Genetic code: SGC3
C;Superfamily: hemolysin homolog yqxc

Query Match 27.7%; Score 352; DB 2; Length 238;
Best Local Similarity 36.8%; Pred. No. 3.2e-23;
Matches 89; Conservative 41; Mismatches 98; Indels 14; Gaps 4;

QY 13 FLOGLAESRBOAKRLIMAGKVTLTNNSTTIPRLKPKGHKYPLESICSLIGVERFVSRA 72
Db 6 YLISLDKNEKEAVGLILSSKKVYNDEFLVLPVLVKQ-----NDIIEVNVKEYVSRGA 59
QY 73 YKLLTALDFFKIDVKSCICLDAGASTGGFTDCLLQHGASKVYIDVGKQGHKLYTNEQ 132
Db 60 YKLLQAIEDFKDKIKVLDIGSGTGGFSAQVINGAKKVVYALDSGTNQLDYKRLIDN 119
QY 133 VINIEGNLRTASKDLIPREVILITDVFSITLTLPSCIRMLKASGIILIKPOPEL 192
Db 120 VVYENTNLKNINPKMFEBQDLCVCDVSPISUKEVFKVLQSLKRNKSLMALIKPOFEA 179
QY 193 YPDKIKKGVVKETSLQYEAIVEKIHFQ--QSELGLIFIGVVPVSIKPKG-NQYLIYL 248
Db 180 PKNLVEKGGVYVNYKHEEIIINKVINFAHDYEFELVQFF----PSKTPGKKAKNIEYISLF 235
QY 249 KK 250
Db 236 RK 237

RESULT 23
E87413
hemolysin A [imported] - Caulobacter crescentus
C;Species: Caulobacter crescentus
C;Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 10-May-2001
C;Accession: E87413
R;Niernan, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg,
B.; Laub, M.T.; Deboy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Kolon
n, J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.M
Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001
A;Title: Complete Genome Sequence of Caulobacter crescentus.
A;Reference number: A87249; MUID:21173698; PMID:11259647
A;Accession: E87413
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-243 <STO>
A;Cross-references: GB:AE005673; NID:G13422669; PIDN:AAK23305.1; GSPDB:GN00148
C;Genetics:
A;Gene: CCL324
C;Superfamily: hemolysin homolog yqxc

Query Match 27.6%; Score 351; DB 2; Length 243;
Best Local Similarity 34.7%; Pred. No. 4e-23;
Matches 85; Conservative 52; Mismatches 100; Indels 8; Gaps 3;

QY 7 RADELVLQGLABREQAARLIMAGKVTLTNNSTTIPRLKPKGHKYPLESICSLIGVER 66
Db 5 RIDQLLVDRGVFDSRAKARAAIEAGRVSVAGRV-----VAKPSEQVDDNAEVAEAAHP 58
QY 67 FYSRGAYKLLTALDFFKIDVKSCICLDAGASTGGFTDCLLQHGASKVYIDVGKQGHK 126
Db 59 WVGKALKLVHALDTWPIAVEGRVAIDVGASTGGFTTEVLLSRGAAFVAVDVGDDQLHPS 118
QY 127 LYTNEOVINIEGNLRTASKDLIPREVILITDVFSITLTLPSCIRMLKASGIITALI 186
Db 119 LRASDRVADLSGVDAASLDDDRIAQPGGLIVSDVFSITLKALPAALHLATRGAEALVALI 178
QY 187 KPQFELYPDKI-KKGVVKETSLQYEAIVEKIHFQCOSELGLIFIGVVPVSIKPKGNQBYL 245
Db 179 KPQFEAGREHVKGKGLVKDPDVIARVEREIVAFLEA-AGWSVNRGLAESPTTGGEGQIERL 237
QY 246 IYLLK 250
Db 238 VWATK 242
```

```
RESULT 24
C82891
hemolysin UU436 [imported] - Ureaplasma urealyticum
C/Species: Ureaplasma urealyticum
C/Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 02-Sep-2000
C/Accession: C82891
R/Glass, J.I.; Lefkowitz, E.J.; Glass, J.S.; Heiner, C.R.; Chen, E.Y.; Casseil, G.H.
submitted to GenBank, February 2000
A/Description: The complete sequence of Ureaplasma urealyticum: Alternate views of a min
A/Reference number: A82870
A/Accession: C82891
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-247 <GLA>
A/Cross-references: GB:AF002140; GB:AF222894; NID:G6899420; PIDN:AAF30848.1; GSPDB:GN001
A/Experimental source: serovar 3; biovar 1
C/Genetics:
A/Gene: hlyA; UU436
C/Superfamily: hemolysin homolog yqxX

Query Match 26.9%; Score 341.5; DB 2; Length 247;
Best Local Similarity 34.6%; Pred. No. 2.8e-22;
Matches 88; Conservative 42; Mismatches 101; Indels 23; Gaps 5;

Qy 7 RADELVFLQGLAESREQAQRLIMAGKVTLTNNSTTTLPLLEKPGHKYPLESICSLIGVER 66
Db 4 RLHLYLKNNTQCETRSKAIDLIKRVRVNNFQVLKPSFLVKQTDTSVEIKN-----DEMY 58

Qy 67 FVSRGAYKLLTALDFFKIDVKSCICLDAGASTGGTDCCLLQHGASKVYVADVGKQLHEK 126
Db 59 FVSKGGYKLLKIEINLIKNFVADLGSSTGGTDCCLQLGAKYIADVGVDLHPS 118

Qy 127 LYTNEQVINIEGVNLRASKDLIPEVDILTIDVSPISLTILPSCIRWLKASGIILALI 186
Db 119 LKSKYKIINVEKTNVXNLDKSYFLEDLDLIYGDLSFISLEQIFFIKKISSPKTLTLLI 178

Qy 187 KPQELFYPDKIK--KGVVKETSLOYEAVEKIHF-----CQSELGLIFIGVPS-VIK 236
Db 179 KPQELGKVASKYKGIQNKLOQLAINKIINLAKNYDFECKH-----LTFTDIFD 230

Qy 237 GPKGNQBYLIYLYKK 250
Db 231 EKKQNEQWIFLQX 244

RESULT 25
F95139
hemolysin A, probable [imported] - Streptococcus pneumoniae (strain TIGR4)
C/Species: Streptococcus pneumoniae
C/Date: 03-Aug-2001 #sequence_revision 03-Aug-2001 #text_change 03-Aug-2001
C/Accession: F95139
R/Tettelin, H.; Nelson, K.E.; Paulsen, I.T.; Eisen, J.A.; Read, T.D.; Peterson, S.; Heid
on, J.D.; Umayam, L.A.; White, O.; Salzberg, S.L.; Lewis, M.R.; Radune, D.; Holtzapfel,
nson, T.; Hickey, E.K.; Holt, I.E.
Science 293, 498-506, 2001
A/Authors: Loftus, B.J.; Yang, F.; Smith, H.O.; Venter, J.C.; Dougherty, B.A.; Morrison,
A/Title: Complete Genome Sequence of a virulent isolate of Streptococcus pneumoniae.
A/Reference number: A95000; MUID:21357209; PMID:11463916
A/Accession: F95139
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-198 <KUR>
A/Cross-references: GB:AB005672; PIDN:AAK75311.1; PID:gl14972684; GSPDB:GN00164; TIGR:SP4
A/Experimental source: strain TIGR4
C/Genetics:
A/Gene: SP1204

Query Match 24.9%; Score 316; DB 2; Length 198;
Best Local Similarity 41.0%; Pred. No. 3.5e-20;
Matches 71; Conservative 33; Mismatches 67; Indels 2; Gaps 2;
```

```
Qy 79 LDRFKIDVKSCICLDAGASTGGTDCCLLQHGASKVYVADVGKQLHEKLYTNEQVINIEG 138
Db 1 MQVFEDLSVDGATTTIDIGASTGGFTDVMQLNSAKLVFAVDVGTNQLAWKLDRPVRVSEQ 60

Qy 139 VNLRASKDLIPEVDILTIDVSPISLTILPSCIRWLKASGIILALIKPQFELYDPDKI- 197
Db 61 FNPRYAEKTDPEQSPFASDVSPISLSLILPALHRVLADQGVVALVVKPQFEAGREQIG 120

Qy 198 KGVVKETSLOYEAVEKIHFQCSSELGLIFIGVVPVSIKPGKGNQBYLIYLYKK 250
Db 121 KNGIIRDAKHQNVLESVTAWA-VEVGFVSLGLDFSPIQGHGNIETFLAYLYKK 172

RESULT 26
F71944
Probable hemolysin - Helicobacter pylori (strain J99)
C/Species: Helicobacter pylori
A/Variety: strain J99
C/Date: 12-Feb-1999 #sequence_revision 12-Feb-1999 #text_change 24-Sep-1999
C/Accession: F71944
R/Alm, R.A.; Ling, L.S.L.; Moir, D.T.; King, B.L.; Brown, B.D.; Doig, P.C.; Smith, D.R.;
; Ives, C.; Gibson, R.; Merberg, D.; Mills, S.D.; Jiang, Q.; Taylor, D.E.; Vovis, G.F.; J
Nature 397, 176-180, 1999
A/Title: Genomic sequence comparison of two unrelated isolates of the human gastric patho
A/Reference number: A71800; MUID:99120557; PMID:9923682
A/Accession: F71944
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-235 <ARN>
A/Cross-references: GB:AF001469; GB:AF001439; NID:94154861; PIDN:AAD05917.1; PID:94154861
A/Experimental source: strain J99
C/Genetics:
A/Gene: hlyA
C/Superfamily: hemolysin homolog yqxX

Query Match 24.2%; Score 307.5; DB 2; Length 235;
Best Local Similarity 35.2%; Pred. No. 2.4e-19;
Matches 87; Conservative 40; Mismatches 105; Indels 15; Gaps 5;

Qy 6 VRADLVFLQGLAESREQAQRLIMAGKVTLTNNSTTTLPLLEKPGHKYPLESICSLIGVE 65
Db 1 MRLDYALFNQHLANSREKAKALVKQLVKNVMVSPSFIVKEDGQ-----IELIAPN 54

Qy 66 FVSRGAYKLLTALDFFKIDVKSCICLDAGASTGGTDCCLLQHGASKVYVADVGKQLHE 125
Db 55 LFVSRAGEKLGAFLEHDFIDFEKVLVDVGASKGFSQVALLKGAKKVLCVDVGKMLDE 114

Qy 126 KLYTNEQVINIEGVNLRASKDLIPEVDILTIDVSPISLTILPSCIRWLKASGIILALI 185
Db 115 SLKNDQRIECVEECDIRGFK--TFEKIDLALCDVSPISLYLCILEAI--LPLSGEFTL 168

Qy 186 IKPOFEL--YDPDKIKGVVKETSLOYEAVEKIHFQCSSELGLIFIGVVPVSIKPGKGNQ 243
Db 169 FKPOFEGVTRTKRKNKGVMDEKAILNALENFNKHLTK-DFQILTITQESLVKKGKNGVE 227

Qy 244 YLIYLYKK 250
Db 228 FFIHFKR 234

RESULT 27
A70163
hemolysin (tlyA) homolog - Lyme disease spirochete
C/Species: Borrelia burgdorferi (lyme disease spirochete)
C/Date: 13-Feb-1998 #sequence_revision 13-Feb-1998 #text_change 24-Sep-1999
C/Accession: A70163
R/Fraser, C.M.; Casjens, S.; Huang, W.M.; Sutton, G.G.; Clayton, R.; Lathigra, R.; White,
son, D.; Peterson, J.; Kerlavage, A.R.; Quackenbush, J.; Salzberg, S.; Hanson, M.; Vugt,
; Bowman, C.; Garland, S.; Fujii, C.; Cotton, M.D.; Horst, K.; Roberts, K.; Hatch, B.
Nature 390, 580-586, 1997
A/Authors: Smith, H.O.; Venter, J.C.
A/Title: Genomic sequence of a Lyme disease spirochete, Borrelia burgdorferi.
```

A;Reference number: A70100; MUID:98065943; PMID:9403685
A;Accession: A70163
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-263 <KLE>
A;Cross-references: GB:AE0001153; GB:AE000783; NID:g2698419; PIDN:AAC66873.1; PID:g269842
A;Experimental source: strain B31
C;Superfamily: hemolysin homolog yqx
Query Match 23.5%; Score 298.5; DB 2; Length 263;
Best Local Similarity 37.4%; Pred. No. 1.7e-18;
Matches 89; Conservative 40; Mismatches 92; Indels 17; Gaps 7;
QY 19 ESREQAKRLIMAGKVTLLPRLERKPGHKYPL-ESICSLIGVERFVSRGAYKLLT 77
DB 19 KTKELMILLKGNIVNSHKENPKLLINKTSKIDLVNTC-----QTFVSRGGYKLL 73
QY 78 ALDFKIDVKSCICLDAGASTGFTDCLLQHGASKVYADVKGQLHEKLYTNQVINIE 137
DB 74 ALKDFEIEVKNKICVDVSGSTGFTDCLLQCGANFVYSIDVGINQLSYKLRIIDPRVKVLE 133
QY 138 GVNLRASK-DLPIEEVDILTIDVSPISLTLLPSCIRWLKASGIIIALIKPOFEL--- 192
DB 134 RTWIFDVTEKIYP---NFAVDVSPRSSISICVNLDKL-SDNFIIVLKPOFEPKSLN 189
QY 193 YPDKIKKGVVKTSLQYEAVERKIHFQCSGLGIFIGVPSVIKPGKNGQYLIYLKK 250
DB 190 LDIKFNGVVGSEVLKTIILQSVIEKFKYKQLQV--KKILKTKGKKGNEFFFLVVK 245
RESULT 28
F64655
hemolysin - Helicobacter pylori (strain 26695)
C;Species: Helicobacter pylori
C;Date: 09-Aug-1997 #sequence_revision 09-Aug-1997 #text_change 24-Sep-1999
C;Accession: F64655
R;Tomb, J.F.; White, O.; Kerlavage, A.R.; Clayton, R.A.; Sutton, G.G.; Fleischmann, R.D.; Peterson, S.; Loftus, B.; Richardson, D.; Dodson, R.; Khalak, H.G.; Glodek, A.; McKenne-son, J.D.; Kelley, J.M.; Cotton, M.D.; Weidman, J.M.; Fujii, C.; Bowman, C.; Watthey, L.; Nature 388, 539-547, 1997
A;Authors: Wallin, E.; Hayes, W.S.; Borodovsky, M.; Karpk, P.D.; Smith, H.O.; Fraser, C.
A;Title: The complete genome sequence of the gastric pathogen Helicobacter pylori.
A;Reference number: A64520; MUID:97394467; PMID:9252185
A;Accession: F64655
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-235 <TOM>
A;Cross-references: GB:AE000615; GB:AE000511; NID:g2314230; PIDN:AAD08129.1; PID:g231423
C;Superfamily: hemolysin homolog yqx
Query Match 22.9%; Score 290.5; DB 2; Length 235;
Best Local Similarity 33.6%; Pred. No. 7.3e-18;
Matches 83; Conservative 41; Mismatches 108; Indels 15; Gaps 5;
QY 6 VRADELVFLQGLAESREQAKRLIMAGKVTLLPRLERKPGHKYPIESTCSLIGVE 65
DB 1 MRLDYALFSGHLYNSREKALVKNQVLYNKMVSPSPFVRENDK-----IELIAEK 54
QY 66 RFVSRGAYKLLTALDFFKIDVKSCICLDAGASTGFTDCLLQHGASKVYADVKGQLHE 125
DB 55 LFVSRAGEKLGAFLETHFDFKGVLDVGASKGSGFSQVALLGAKRVLCVDVGKQOLDE 114
QY 126 KLYTNQVINIEGVNLRASKDLPIEEVDILTIDVSPISLTLLPSCIRWLKASGIIIAL 185
DB 115 SLAQDKRIECYEECDIRGPK--TPETIDIALCDVSPISLYILEAI--LPLSDEFLT 168
QY 186 IKQOFELYP--DKIKGVVKTSLQYEAVERKIHFQCSGLGIFIGVPSVIKPGKNGQE 243
DB 169 FKQFQFVGRGIRKKNKGWVDEKAILNALNLENFKNHLTK-DFQILKIQESLVKKGNGVE 227
QY 244 YLIYLKK 250
DB 228 FFIHFR 234

RESULT 29

C81406
probable hemecolysin Cj0588 [imported] - Campylobacter jejuni (strain NCTC 11168)
C;Species: Campylobacter jejuni
C;Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 03-Jun-2002
C;Accession: C81406
R;Parkhill, J.; Wren, B.W.; Mungall, K.; Ketley, J.M.; Churcher, C.; Basham, D.; Chiller, C.W.; Quail, M.; Rajandream, M.A.; Rutherford, K.M.; VanVliet, A.; Whitehead, S.; Barrel-
Nature 403, 665-668, 2000
A;Title: The genome sequence of the food-borne pathogen Campylobacter jejuni reveals hylE
A;Reference number: A81250; MUID:20150912; PMID:10688204
A;Accession: C81406
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-253 <PAR>
A;Cross-references: GB:AL139075; GB:AL111168; NID:g6967817; PIDN:CAB75224.1; PID:g696805
A;Experimental source: serotype O2, strain NCTC 11168
C;Genetics:
A;Gene: tlyA; Cj0588
C;Superfamily: hemolysin homolog yqx
Query Match 21.2%; Score 270; DB 2; Length 253;
Best Local Similarity 38.3%; Pred. No. 4.9e-16;
Matches 74; Conservative 31; Mismatches 73; Indels 12; Gaps 5;
QY 65 ERFVSRGAYKLLTALDFFKIDVKSCICLDAGASTGFTDCLLQHGASKVYADVKGQLH 124
DB 71 EIVSRAALKKXFELENDIEIKHKNCLDIGSGTGFVQILLENQALKITALDVGSNQLH 130
QY 125 EKLYTNQVINIEGVNLRASKDLPIEEVDILTIDVSPISLTLLPSCIRWLKASGI--I 182
DB 131 PSRLVNEKIILHENTDLRAFKSE---EKFLVTCDSVFSISLINLL----YYIDNLAKEI 183
QY 183 IALIKPOFELYP--KIKGVVKTSLQYEAVERKIHFQCSGLGIFIGVPSVIKPGK 240
DB 184 ILFKPQFQFVGRGIRKKNKGWVDEKAILKARMDFFKAC-AKIGLLKNTQKSSIKGKG 242
QY 241 NOEYLIYLKK 250
DB 243 NVEYFYIYK 252
RESULT 30
H69522
conserved hypothetical protein AF2184 - Archaeoglobus fulgidus
C;Species: Archaeoglobus fulgidus
C;Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 22-Oct-1999
C;Accession: H69522
R;Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dodson, J.; Fleischmann, R.D.; Quackenbush, J.; Lee, N.H.; Sutton, G.G.; Gill, S.; Kirkness, E.F.; Glodek, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Weidman, J.F.; McDonald, L.; Nature 390, 364-370, 1997
A;Authors: Utterback, T.; Cotton, M.D.; Spriggs, T.; Artiach, P.; Kaine, B.P.; Sykes, S.; Smith, H.O.; Woese, C.R.; Venter, J.C.
A;Title: The complete genome sequence of the hyperthermophilic, sulfate-reducing archae
A;Reference number: A69250; MUID:98049343; PMID:9389475
A;Accession: H69522
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-208 <KLE>
A;Cross-references: GB:AE000954; GB:AE000782; NID:g2699277; PIDN:AAB89068.1; PID:g264883
Query Match 9.2%; Score 117; DB 2; Length 208;
Best Local Similarity 26.3%; Pred. No. 0.0084;
Matches 50; Conservative 31; Mismatches 87; Indels 22; Gaps 7;
QY 6 VRADELVFLQGLAESREQAKRLIMAGKVTLLPRLERKPGHKYPIESTCSLIGVE 65
DB 1 MRLDILLVRGFFSSRAKEAIKKGFLVDGKKVT-----KPSAEVDFEAEIKVLQPE 54
QY 66 RFVSRGAYKLLTALDFFKIDVKSCICLDAGASTGFTDCLLQHGASKVYADVKGQLHE 125

```
Db      55 R--PRGYWKLKDEIDHFNLFSGNEVVDLGGSSAGFL-LYASEKAEVGYIEYSR-EPEE 110
      126 KLYTNEQVINTGVNLTASKDLIPEEVDIL-TIDVSFISLTILPSCIR-----WLK 177
      111 GL-----REIERQPNKVFIADAFDTLSLPELDLIINDLTLPSSFRAMRRLPLLK 166
      178 ASGIIIALIK 187
      167 KGGIVLFVHK 176

RESULT 31
H64456
Hypothetical protein MJ1257 - Methanococcus jannaschii
C:Species: Methanococcus jannaschii
C>Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 21-Jul-2000
C:Accession: H64456
R:Bult, C.J.; White, O.; Olsen, G.J.; Zhou, L.; Fleischmann, R.D.; Sutton, G.G.; Blake,
; Reich, C.I.; Overbeek, R.; Kirkness, E.F.; Weinstock, K.G.; Merrick, J.M.; Glodek, A.;
rson, J.D.; Sadow, P.W.; Hanna, M.C.; Cotton, M.D.; Roberts, K.M.; Hurst, M.A.
Science 273, 1058-1073, 1996
A:Authors: Kaine, B.P.; Borodovsky, M.; Klenk, H.P.; Fraser, C.M.; Smith, H.O.; Woese, C
A>Title: Complete genome sequence of the methanogenic archaeon, Methanococcus jannaschii
A:Reference number: A64300; MUID:96337999; PMID:8688087
A:Accession: H64456
A>Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-349 <BUL>
A:Cross-references: GB:U67566; GB:L77117; NID:g1591887; PIDN:AAB99268.1; PID:g1591891; T
C:Genetics:
A:Map position: FOR1199545-1200594

Query Match      8.9%; Score 112.5; DB 2; Length 349;
Best Local Similarity 28.1%; Pred. No. 0.039; Mismatches 28; Indels 31; Gaps 8;
Matches 45; Conservative 28;

Qy      64 VERFVSRGAYKLLTALDFFKIDVKSCIC-LDAGASTGGFTDCLLQHGASKVYVADVKGQ 122
      182 IERPLNSRERKQWELMEKFFIPENINCVDIGSGPGWAK-MLSKAKKYYAIDTGE-- 238
      123 LHEKLYNEQVINTEGVNLTASKDL---IPEEVDLTIDVSFI---SLTILPSCIRWL 176
      239 --LKAIA---NIIHKRAENVDFEKDINEEDLIINDLTNLYPDESIFLTL-KFAKHL 291
      177 KASGIIIALIKPQPELYPDKIKGVKETSQYEAVERKII 216
      292 KTNGYIIHTLK-----ARNLTKKSDLEKVL 317

RESULT 32
C86784
Pseudouridine synthase [imported] - Lactococcus lactis subsp. lactis (strain IL1403)
C:Species: Lactococcus lactis subsp. lactis
C>Date: 23-Mar-2001 #sequence_revision 23-Mar-2001 #text_change 03-Aug-2001
C:Accession: C86784
R:Botolin, A.; Wincker, P.; Mauger, S.; Jaillon, O.; Malarne, K.; Weissenbach, J.; Ehrli
Genome Res. 11, 731-753, 2001
A>Title: The complete genome sequence of the lactic acid bacterium Lactococcus lactis ss
A:Reference number: A86625; MUID:2123186; PMID:11337471
A:Accession: C86784
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-257 <STO>
A:Cross-references: GB:AE005176; PID:g12724251; PIDN:AAK05373.1; GSPDB:GN00146
A:Experimental source: strain IL1403
C:Genetics:
A:Gene: rluB
C:Superfamily: conserved hypothetical protein HU1243

Query Match      8.4%; Score 106.5; DB 2; Length 257;
Best Local Similarity 28.4%; Pred. No. 0.039; Mismatches 29; Indels 43; Gaps 11;
Matches 55; Conservative 29;
```

```
Qy      5 KRADELVFLQGLAESREQAERLIMAGKVTLTNSTTIPLRIEKPGHKYPLESICSLIGV 64
      18 KMRINKYLAHAGVA-SRKAEELIAGKVTVNN-----VP--MTNLGYQVSTGDIVEVNGV 70
      65 ERFVSRGAYKLL-----TALDFFKIDVKSCIC-----LDAGAS----- 97
      71 AVYNEEPVYLLNKPGRGYISSVDDKGRQTWDLMP-QVKERIYPVGRLDWDTSGLLLLT 129
      98 -TGFTDCLL--QHGASKVYVADVKGQLH-EKLYTNEQVINIEGVNLTASKDLIPEEV 153
      130 NDGEFTNMTHPRHGVDKVVYAKV-EGQANKENLRPLTGLMTIEGKKVSPARYEIIKQE- 187
      154 DILTIDVSFISLTL 167
      188 --KTNKHSIVSLTI 199

RESULT 33
B81917
Probable chaperone protein NMA0736 [imported] - Neisseria meningitidis (strain Z2491 serc
C:Species: Neisseria meningitidis
C>Date: 05-May-2000 #sequence_revision 05-May-2000 #text_change 02-Feb-2001
C:Accession: B81917
R:Parkhill, J.; Achtman, M.; James, K.D.; Bentley, S.D.; Churcher, C.; Klee, S.R.; Morell
; Holroyd, S.; Jagels, K.; Leather, S.; Moule, S.; Mungall, K.; Quail, M.A.; Rajandream,
Nature 404, 502-506, 2000
A>Title: Complete DNA sequence of a serogroup A strain of Neisseria meningitidis Z2491.
A:Reference number: A81775; MUID:20222556; PMID:10761919
A:Accession: B81917
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-642 <PAR>
A:Cross-references: GB:AL162754; GB:AL157959; NID:g7379424; PIDN:CAB84020.1; PID:g737945f
A:Experimental source: serogroup A, strain Z2491
C:Genetics:
A:Gene: dnaK; NMA0736
C:Superfamily: heat shock protein 70

Query Match      7.4%; Score 93.5; DB 2; Length 642;
Best Local Similarity 23.7%; Pred. No. 3.8;
Matches 62; Conservative 32; Mismatches 89; Indels 79; Gaps 12;

Qy      10 ELVFLOGLAESREQAERLIMAGKVTLTN-----NSTTIPLRIEKPGHKYPLESICSLIGV 64
      256 DVMAIQRLKKAERAKIELSSGQOTEINLPYTNDAFGPKHLAMKITRAKFESL-----V 310
      65 ERFVSRGAYKLLTALDFFKIDVKSCICLDAGASTGGFTDCLLQHGASKVYVADVKGQLH 124
      311 EDLIARSIEPRTALK-----DAGLSTGDDIDVILVGQSRMPKV----- 350
      125 EKLYTNEQVINIEGVNLTASKDLIPEEV-----DILTIDVSFISLTL 167
      351 -----QEAVERDFFGKEPR---KDVNPDEAVAVGAAIQGEVLSSGSRSDVLLIDVTPLSLGI 402
      168 ILPSCIRWLKASGIIIALIKPQPELYPDKIKGVKETSQYEAVERKIIHFQSEF----- 222
      403 -----ETWGGVMTKLIQKNTTI-PTKASQ-VFSTAEDNQSAV--TIHVLQGERERAS 450
      223 ----LGLIFIGVVPSEVIKGPKG 240
      451 ANKSLGQFNLG---DIAPAPRG 469

RESULT 34
AF1302
RNA-binding Sun protein homolog lmo1822 [imported] - Listeria monocytogenes (strain EGD-e
C:Species: Listeria monocytogenes
C>Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 01-Mar-2002
C:Accession: AF1302
R:Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloecker,
; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi, H.;
D.; Jones, L.M.; Karst, U.
```


Db |: |:|
 459 PA----PRG 463

Search completed: July 7, 2004, 16:57:35
Job time : 42.5149 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 7, 2004, 16:47:59 ; Search time 16.6777 Seconds
(without alignments)
783.656 Million cell updates/sec

Title: US-10-009-919A-1

Perfect score: 1271

Sequence: 1 MAKHKVRADELVLFLGLAES.....PSVIKPGKQNEYLIYLKKR 251

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_42:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	418	32.9	269	1	YQXC_BACSU
2	383.5	30.2	240	1	HLYA_TREHY
3	112.5	8.9	349	1	YCS7_METJA
4	93.5	7.4	642	1	DNAK_NEIMA
5	91	7.2	747	1	V222_FOWPV
6	89	7.0	328	1	YD70_METJA
7	88.5	7.0	627	1	DNAK_GUTHA
8	88	6.9	191	1	YK51_ARCFU
9	87.5	6.9	494	1	MURE_PASMU
10	87.5	6.9	642	1	DNAK_NEIMB
11	87	6.8	640	1	DNAK_PORGI
12	87	6.8	644	1	DNAK_NITEU
13	87	6.8	1853	1	Y99104_mus musculus
14	86	6.8	501	1	Y890_THEMEA
15	86	6.8	1828	1	Q9CYF3_rattus norv
16	86	6.8	1855	1	MY5A_HUMAN
17	85.5	6.7	200	1	R54_OEIH
18	85.5	6.7	618	1	DNAK_DEIPR
19	84.5	6.6	200	1	R54_BACHD
20	84	6.6	203	1	R54_STRMU
21	84	6.6	217	1	GRPE_MYCGE
22	84	6.6	334	1	ANM4_HUMAN
23	83	6.5	203	1	R54_STRPN
24	83	6.5	203	1	R54_STRPY
25	83	6.5	355	1	KPR4_YEAST
26	83	6.5	464	1	GFR2_HUMAN
27	83	6.5	466	1	GPT_LEIME
28	83	6.5	631	1	DNAK_PASHA
29	83	6.5	3214	1	BPA1_HUMAN
30	82.5	6.5	199	1	R54_BACSU
31	82.5	6.5	460	1	TRME_THETN
32	82.5	6.5	627	1	DNAK_DEIRA
33	82.5	6.5	637	1	HS70_SCHWA

34	82	6.5	318	1	KPR2_YEAST
35	82	6.5	609	1	DNAK_ENTFA
36	81.5	6.4	203	1	RS4_IACIA
37	81.5	6.4	309	1	PRM4_THETN
38	81.5	6.4	635	1	DNAK_TREPA
39	81.5	6.4	638	1	DNAK_PRESM
40	81	6.4	199	1	RS4_BACST
41	81	6.4	462	1	TRME_STRAG
42	81	6.4	463	1	GFR2_MOUSE
43	81	6.4	664	1	AROB_FUSNN
44	80.5	6.3	320	1	OXE2_RAT
45	80.5	6.3	416	1	MURA_BUCAI

ALIGNMENTS

RESULT 1
YQXC_BACSU STANDARD; PRT; 269 AA.
AC P19672:
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Hypothetical protein yqxc.
GN YQXC OR YQIF OR BSU24260.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=168 / JH642;
RX MEDLINE=97124195; PubMed=8969508;
RA Mizuno M., Masuda S., Takemaru K.-I., Hosono S., Sato T., Takeuchi M., Kobayashi Y.;
RT "Systematic sequencing of the 283 kb 210 degrees-232 degrees region of the Bacillus subtilis genome containing the skin element and many sporulation genes.";
RL Microbiology 142:3103-3111(1996).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=98044033; PubMed=9384377;
RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G., Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S., Borriess R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S., Brouillet S., Bruschi C.V., Connerton I.F., Cummings N.J., Daniel R.A., Denizot F., Devane K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T., Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D., Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N., Ghim S.-Y., Glaser P., Goffeau A., Golightly E.J., Grandi G., Giuseppe G., Guy B.-J., Haga K., Halech J., Harwood C.R., Henaut A., Hilbert H., Holstappel S., Hosono S., Hullo M.F., Itaya M., Jones L., Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C., Kobayashi Y., Koetter P., Koningsstein G., Krogh S., Kumano M., Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V., Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigue C., Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M., Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H., Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M., Presecan E., Puig P., Purnelle B., Rapoport G., Rey M., Reynolds S., Rieger M., Rivolta C., Roche E., Roche B., Rose M., Sadaie Y., Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F., Soldo B., Sekiguchi J., Sekowska A., Seror S.J., Seror P., Shin B.S., Soldo B., Sorokin A., Taccini E., Takagi T., Takahashi H., Takemaru K., Takeuchi M., Tanakoshi A., Tanaka T., Terpstra P., Tognotti A., Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A., Viari A., Wambut R., Wedler E., Wedler H., Weitzenegger T., Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K., Yoshida K., Yoshikawa H.F., Zumstein E., Yoshikawa H., Danchin A.;
RT "The complete genome sequence of the Gram-positive bacterium Bacillus subtilis.";

```
RL Nature 390:249-256 (1997).
RN [3]
RP SEQUENCE OF 205-269 FROM N.A.
RC STRAIN=168 / EMG50;
RX MEDLINE=9006783; PubMed=2507400;
RA North A.K., Smith M.C.M., Baumberg S.;
RT "Nucleotide sequence of a Bacillus subtilis arginine regulatory gene
RT and homology of its product to the Escherichia coli arginine
RT repressor.";
RL Gene 80:29-38 (1989).
CC -!- SIMILARITY: Contains 1 S4 RNA-binding domain.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; D84432; BA012577.1; -.
CC EMBL; Z99116; CAB14357.1; -.
CC EMBL; M27869; AAA22207.1; -.
CC DR PIR; G69967; G69967.
CC Subtilist; BG10308; yqxC.
CC InterPro; IPR002877; RrmJ_FtsJ.
CC InterPro; IPR002942; S4.
CC InterPro; IPR004538; Tly.
CC Pfam; PF01728; FtsJ; 1.
CC Pfam; PF01479; S4; 1.
CC SMART; SM00363; S4; 1.
CC TIGRFAMS; TIGR00478; tly; 1.
CC DR PROSITE; PSS0889; S4; 1.
CC KW Hypothetical protein; RNA-binding; Complete proteome.
FT DOMAIN 6 67 S4 RNA-BINDING.
FT CONFLICT 205 210 HADVLK -> CMLIAQ (IN REF. 3).
SQ SEQUENCE 269 AA; 29652 MW; 77802BC5F8AC97F6 CRC64;

Query Match 32.9%; Score 418; DB 1; Length 269;
Best Local Similarity 43.4%; Pred. No. 1.1e-27;
Matches 108; Conservative 38; Mismatches 93; Indels 10; Gaps 5;

QY 1 MAHKVRADELVFLQGLAESREQAKRLIMAGKVTLTNNSTTIPLEKPKGKYPLESICS 60
Db 1 MTSKGRDLVLLVERGLAETREKAKRAIMAGIYVSNEN-----RLDKPGEKIDRLPLT 54
QY 61 LIGVE-RFVSRGAYKLLTALDFPKIDVKSCICLDAGASTGGTDCLLQHGASKVYALDVG 119
Db 55 VKGNPLRYVSRLGKLEKALKKEPVSVKDKIMIDIGSSTGGTDCALONGAKOSYADVVG 114
QY 120 KQGLHEKLYTNEQVINIGVNLRTAS-KOLIPREVDTLTDVFSISLTILPSCIRWLKA 178
Db 115 YNLQAWKLQRDERVVVMTNTFRYATPADPTKGMPEFATIDVFSISLRLILPVLRLTLLP 174
QY 179 SGIIILIKPEPELYDPKI-KKGVKETSLOYEAVEKIIHFCOSELGLIFGVVPSVIKG 237
Db 175 GSDCMALVPQPEAGRESVGKGIIVDPKVHADVLKRMISFSAAB-GYICKGLSFSFPTG 233
QY 238 PKGNQYILI 246
Db 234 GDGNIEYLL 242

RESULT 2
HLVA TREHY
ID HLVA TREHY STANDARD; PRT; 240 AA.
AC Q06803;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Hemolysin A.
GN TLY OR TLY.
OS Treponema hyodysenteriae (Serpulina hyodysenteriae).
```

```
OC Bacteria; Spirochaetes; Spirochaetales; Brachyspiraceae; Brachyspira.
OX NCBI_TaxID=159;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=B204;
RX MEDLINE=92112315; PubMed=1730486;
RA Muir S., Koopman M.B.H., Libby S.J., Joens L.A., Heffron F.,
RA Kusters J.G.;
RT "Cloning and expression of a Serpula (Treponema) hyodysenteriae
RT hemolysin gene.";
RL Infect. Immun. 60:529-535 (1992).
CC -!- FUNCTION: Bacterial hemolysins are exotoxins that attack blood
CC cell membranes and cause cell rupture by mechanisms not clearly
CC defined.
CC -!- INDUCTION: By sodium ribonuclease.
CC -!- SIMILARITY: Contains 1 S4 RNA-binding domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X61684; CAA43859.1; -.
CC EMBL; A17012; CAA01317.1; -.
CC DR PIR; A43863; A43863.
CC InterPro; IPR002877; RrmJ_FtsJ.
CC InterPro; IPR002942; S4.
CC InterPro; IPR004538; Tly.
CC Pfam; PF01728; FtsJ; 1.
CC Pfam; PF01479; S4; 1.
CC SMART; SM00363; S4; 1.
CC TIGRFAMS; TIGR00478; tly; 1.
CC DR PROSITE; PSS0889; S4; 1.
CC KW Hemolysis; Toxin; RNA-binding.
FT DOMAIN 1 61 S4 RNA-BINDING.
FT CONFLICT 240 240 26881 MW; E718A0A21DF3CCB7 CRC64;
SQ SEQUENCE 240 AA; 26881 MW; E718A0A21DF3CCB7 CRC64;

Query Match 30.2%; Score 383.5; DB 1; Length 240;
Best Local Similarity 38.6%; Pred. No. 7.1e-25;
Matches 96; Conservative 41; Mismatches 93; Indels 19; Gaps 5;

QY 6 VRADELVFLQGLAESREQAKRLIMAGKVTLTNNSTTIPLEKPKGKYPLESICSIGVE 65
Db 1 MRLDEVVHSEGVSTESKQADIIAGCVFVG-----VKVTSKAHKIKDNDIEVVQNI 54
QY 66 RFVSRGAYKLLTALDFPKIDVKSCICLDAGASTGGTDCLLQHGASKVYALDVGKQLHE 125
Db 55 KYVSRAAGEKLEKAFVEFGISVENKICLDIGASTGGTDCLLQHGAKKVVYALDVGHNQLVY 114
QY 126 KLYTNEQVINIGVNLRTASKDLIPREV-DILTIDVYSFISLTILPSCIRWLKASGIIIA 184
Db 115 KLNRNRRVVSIEDFNKAKINKEMFNDEIFSVISVDSFISITKIAIIPKELNNLEFWVT 174
QY 185 LIKPEPELYDPKIKKG-----VVKETSLOYEAVEKIIHFCOSELGLIFGVVPSVIKGP 238
Db 175 LIKPEPEAEERGVDVSGGIIRDDILREKILN-NAISKII-----DCGFKVNRVTSPIKGA 238
QY 239 KGNQYILI 247
Db 229 KGNIEYLAH 237

RESULT 3
YC57 METJA
ID YC57 METJA STANDARD; PRT; 349 AA.
AC Q58654;
DT 15-MAR-2004 (Rel. 43, Created)
DT 15-MAR-2004 (Rel. 43, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Hypothetical protein MJ1257.
```



```
OC Viruses; dsDNA viruses, no RNA stage; Poxviridae; Chordopoxvirinae;
OC Avipoxvirus.
OX NCBI_TaxID=10261;
RN [1]
RX SEQUENCE FROM N.A.
RX MEDLINE=20193820; PubMed=10729156;
RA Afonso C.L., Tulman E.R., Lu Z., Zsak L., Kutish G.F., Rock D.L.;
RT "The genome of fowlpox virus.";
RL J. Virol. 74:3815-3831 (2000).
CC -!- SIMILARITY: Contains 14 ANK repeats.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL; AF198100; AAF44566.1; -.
DR HSP; Q00420; 1AWC.
DR InterPro; IPR002110; ANK.
DR InterPro; IPR001810; F-box.
DR Pfam; PF000646; F-box; 1.
DR PRINTS; PR01415; ANKYRIN.
DR SMART; SM00248; ANK; 15.
DR PROSITE; PS50088; ANK_REPEAT; 7.
DR PROSITE; PS50297; ANK_REPEAT_REGION; 1.
KW Hypothetical protein; Repeat; ANK repeat.
FT REPEAT 38 67 ANK 1.
FT REPEAT 103 132 ANK 2.
FT REPEAT 136 165 ANK 3.
FT REPEAT 169 198 ANK 4.
FT REPEAT 202 231 ANK 5.
FT REPEAT 234 263 ANK 6.
FT REPEAT 294 323 ANK 7.
FT REPEAT 328 357 ANK 8.
FT REPEAT 361 393 ANK 9.
FT REPEAT 397 426 ANK 10.
FT REPEAT 430 460 ANK 11.
FT REPEAT 464 493 ANK 12.
FT REPEAT 495 524 ANK 13.
FT REPEAT 529 559 ANK 14.
SQ SEQUENCE 747 AA; 85303 MW; 55F90AF2855C3D28 CRC64;

Query Match 7.2%; Score 91; DB 1; Length 747;
Best Local Similarity 23.5%; Pred. No. 5.7; Mismatches 84; Gaps 12;
Matches 59; Conservative 31;

QY 22 EQAKRLIMAGKVTLTNNSTTI-PIRLKPKGHKYPLESICSLIGVERFVSRGAYKLLTALD 80
Db 118 ELAKMLIQNALVNMNMKNITPLHIAS-----SSGYKMW----- 153

QY 81 FFKIDVKSCTCLDAGASTGFTDCLLOHGASKVAIDVGQOLHEKLYTNEQVINI----- 136
Db 154 -----ELLHLGANTNALTSGYTSLSHYSSND--LNISELL 189

QY 137 --EGVNLRTASKDLIPEEVDILTDFSIILTPSCIRWLKASGI-IIALIKPOPELY 193
Db 190 IENGTVNVANKD-----SITALLIAVIMSIDL-----VRLLDKGADTNAIGLERFLY 240

QY 194 PDKIKK--GVVK--ETSLQYAEVEKLIHFQSEL-----GLIFIG-----VVPSV 234
Db 241 VTETKQNNILKYLTNTNNTVNTVMTINEYIASELYDWNRSATSKLMFRSCFPCFVPT 300

QY 235 IKGPKGNQEVYL 245
Db 301 LATRKSGSKELL 311

RESULT 6
YD70_METUA
```

```
ID YD70_METUA STANDARD; PRT; 328 AA.
AC Q58765;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Hypothetical UPF0285 protein M31370.
GN M31370.
OS Methanococcus jannaschii.
OC Archaea; Euryarchaeota; Methanococci; Methanococcales;
OC Methanocaldococcaceae; Methanocaldococcus.
OX NCBI_TaxID=2190;
RN [1]
RX SEQUENCE FROM N.A.
RX STRAIN=JAL-1 / DSM 2661 / ATCC 43067;
RX MEDLINE=96337999; PubMed=8688087;
RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
RA Sutton G.G., Blake J.A., FitzGerald L.M., Clayton R.A., Gocayne J.D.,
RA Kerlavage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.I.,
RA Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,
RA Scott J.L., Geoghagen N.S.M., Weidman J.F., Sadow P.W., Nguyen D.,
RA Utterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,
RA Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
RT "Complete genome sequence of the methanogenic archaeon, Methanococcus
RT jannaschii.";
RL Science 273:1058-1073 (1996).
CC -!- SIMILARITY: Belongs to the UPF0285 family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL; U67577; AAB99387.1; -.
DR PIR; A64471; A64471.
DR TIGR; M31370; -.
DR HAMAP; MF_01087; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 328 AA; 35979 MW; AC9680820123FACF CRC64;

Query Match 7.0%; Score 89; DB 1; Length 328;
Best Local Similarity 27.1%; Pred. No. 3.4; Mismatches 22; Indels 52; Gaps 12;
Matches 54; Conservative 22;

QY 72 AYKLLTALDPFKIDVKS---CICLDAGASTGGTDCI---LQHG---ASKVVAIDVGKG 121
Db 141 AYKLFQFNDFVLSDISNTVTLLIKDKGFGGACIGALMLKGPIDLEWIRDIDAGKI 200

QY 122 QLHE-----KLY-----TNEQVIN--IEGVNLTASKDLIPEEVDILTIDVSFI 163
Db 201 TANEAFSGAGAKIAKLYKGVNTKEEIIINNYENDENCLAVDSL-----ILSVSMEN 254

QY 164 SLTILPSCIRWLKASGIITAIKPOPELYPDKIKGVKETSLOYEAVEKLIHFQSEL 223
Db 255 SLPLLDKKNKRVVLGSGICTLRNP-IDI-PKRIK-----EFVEAKIFVLVYGES 301

QY 224 GLIFIG---VVPVSKGPK 239
Db 302 GA--IGGALLAEIDLKGR 318

RESULT 7
DNAA_GUITH STANDARD; PRT; 627 AA.
AC P29215; 078509;
DT 01-DEC-1992 (Rel. 24, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Chapterone protein dnaK (Heat shock protein 70) (Heat shock 70 kDa
DE protein) (HSP70).
```



```
RESULT 9
MURE_PASMU
ID MURE_PASMU STANDARD; PRT; 494 AA.
AC P57815;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE UDP-N-acetylmuramoyl-L-glutamate-2,6-diaminopimelate ligase
DE (EC 6.3.2.13) (UDP-N-acetylmuramyl-tripeptide synthetase) (Meso-
DE diaminopimelate-adding enzyme) (UDP-MurNAc-tripeptide synthetase).
DE MURE OR PM0137.
GN Pasteurella multocida.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
OC Pasteurellaceae; Pasteurella.
OX NCBI_TaxID=747;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=Pm70;
RX MEDLINE=21145866; PubMed=11248100;
RA May B.J., Zhang Q., Li L.L., Paustian M.L., Whittam T.S., Kapur V.;
RA "Complete genomic sequence of Pasteurella multocida Pm70.";
RA Proc. Natl. Acad. Sci. U.S.A. 98:3460-3465(2001).
RL
CC -!- FUNCTION: Cell wall formation. Diaminopimelic acid adding enzyme
CC (By similarity).
CC
CC -!- CATALYTIC ACTIVITY: ATP + UDP-N-acetylmuramoyl-L-alanyl-D-
CC glutamate + meso-2,6-diaminoheptanedioate = ADP + phosphate + UDP-
CC N-acetylmuramoyl-L-alanyl-D-gamma-glutamyl-meso-2,6-diamino-
CC heptanedioate.
CC
CC -!- PATHWAY: Peptidoglycan biosynthesis.
CC
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (Probable).
CC
CC -!- SIMILARITY: Belongs to the murCDEF family.
CC
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AE006048; AAK02221.1; -.
CC DR HAMAP; MF 00208; -.
CC DR InterPro; IPR000713; Mur ligase.
CC DR InterPro; IPR004101; Mur ligase_C.
CC DR InterPro; IPR005761; MurE.
CC DR Pfam; PF01225; Mur_ligase; 1.
CC DR Pfam; PF02875; Mur_ligase_C; 1.
CC DR TIGRfam; TIGR01085; murE; 1.
CC DR Peptidoglycan synthesis; Cell wall; Cell division; Ligase;
CC ATP-binding; Complete proteome.
CC NP BIND 113 119 ATP (POTENTIAL).
CC FT SEQUENCE 494 AA; 54171 MW; 6C57379960A001D CRC64;
CC
CC Query Match 6.9%; Score 87.5; DB 1; Length 494;
CC Best local Similarity 24.0%; Pred. No. 7.1;
CC Matches 46; Conservative 25; Mismatches 64; Indels 57; Gaps 9;
CC
CC QY 26 RLIMAGVVTNNSTIPLRLEK-----PGHKYPLESICSLIG----- 63
CC
CC Db 105 KTLVVG-VTNGNKTITQSLLAQTLLGHR-----SAVNGTIGNGLLQGVKEATNTTGSV 160
CC
CC QY 64 -----VERFVSRGAYKLLTALDFKIDVKS-----CICLDAGASTGGFTDCLLQHG 109
CC
CC Db 161 EVQASLADFVKRGA-----DFAAIEVSSHGLVQHRVEALAFDVAIFTLNSRDHLDYHQ 213
CC
CC QY 110 ASKVATDVGGQQLHKEKLYTNEQVINEGVNLTASKDLIPEEVDIITIDVSFSLILIL 169
CC
CC Db 214 SMENYAL-AKKRLFTELNSRHIINADD-----SVGQTLQEQPNVAVASQ-----TDFX 263
CC
CC QY 170 PSCIRWLKASGI 181
CC
CC Db 264 PQQARWLKATAI 275
```

```
RESULT 10
DNAK_NEIMB
ID DNAK_NEIMB STANDARD; PRT; 642 AA.
AC Q9K0N4;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Chaperone protein dnaK (Heat shock protein 70) (Heat shock 70 kDa
DE protein) (HSP70).
DE DNK OR NMB0554.
GN Neisseria meningitidis (serogroup B).
OS Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;
OC Neisseriaceae; Neisseria.
OX NCBI_TaxID=491;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=MC58 / Serogroup B;
RX MEDLINE=20175755; PubMed=10710307;
RA Tettelin H., Saunders N.J., Heidelberg J., Jeffries A.C., Nelson K.E.,
RA Eisen J.A., Ketchum K.A., Hood D.W., Peden J.F., Dodson R.J.,
RA Nelson W.C., Gwinn M.L., DeBoy R., Peterson J.D., Hickey E.K.,
RA Haft D.H., Salzberg S.L., White O., Fleischmann R.D., Dougherty B.A.,
RA Mason T., Ciecko A., Parksey D.S., Blair E., Cittone H., Clark E.B.,
RA Cotton M.D., Uutterback T.R., Khouri H., Qin H., Vamathevan J.,
RA Gill J., Scariato V., Maignani V., Pizzo M., Grandi G., Sun L.,
RA Smith H.O., Fraser C.M., Moxon E.R., Rappuoli R., Venter J.C.;
RA "Complete genome sequence of Neisseria meningitidis serogroup B strain
RT MC58.";
RT Science 287:1809-1815(2000).
RL
CC -!- FUNCTION: Acts as a chaperone (By similarity).
CC
CC -!- INDUCTION: By stress conditions e.g. heat shock (By similarity).
CC
CC -!- SIMILARITY: Belongs to the heat shock protein 70 family.
CC
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AE002411; AAF40982.1; -.
CC DR PIR; H81185; H81185.
CC DR HSP; P04475; 1DG4.
CC DR TIGR; NMB0554; -.
CC DR HAMAP; MF 00332; -.
CC DR InterPro; IPR001033; Hsp70.
CC DR Pfam; PF00012; HSP70; 1.
CC DR PRINTS; PR00301; HEATSHOCK70.
CC DR ProDom; PD000089; Hsp70; 1.
CC DR PROSITE; PS00297; HSP70_1; 1.
CC DR PROSITE; PS00329; HSP70_2; 1.
CC DR PROSITE; PS01036; HSP70_3; 1.
CC DR Chaperone; ATP-binding; Heat shock; Phosphorylation;
CC Complete proteome.
CC MOD RES 200 200 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
CC FT SEQUENCE 642 AA; 68791 MW; 235696C763BD5805 CRC64;
CC
CC Query Match 6.9%; Score 87.5; DB 1; Length 642;
CC Best local Similarity 23.7%; Pred. No. 9.5;
CC Matches 62; Conservative 35; Mismatches 86; Indels 79; Gaps 13;
CC
CC QY 10 ELVFLQGLABSRQAKELIMAGKVTLTN-----NSTTIPLRLPKPGHKYPLESICSLIG 64
CC
CC Db 256 DVMAQLQLEAEAKAKIELSSGQQTLEINPYITWDATGPKHAKMITRAKFESL-----V 310
CC
CC QY 65 ERFVSRGAYKLLTALDFKIDVKSICLDAGASTGGFTDCLQHGASKVVAIDVGKQLH 124
CC
CC Db 311 EDLITR-----SIEPCKIALK-----DAGLSTGDDIVLVGGQSRMPKV----- 350
CC
CC QY 125 EKYLTNEQVINIEGVNLTASKDLIPEEV-----DILTIDVSFISLTL 167
```



```

Db 351 -----QEAVKAFPGKEPR---KDVNPDEAVAVGAATQGEVLGGGRSDVLLLDVTPLSLGI 402
QY 168 ILPSCIRWLKASGIIIALIKPQFELYPDKIKKGVKETSLOVEAVEKIHFQOSE-----222
Db 403 -----ETMGVWTKLIQKNNTI-PTKASQ-VFSTAEDNOSAV--TIHVLQGERERAS 450
QY 223 ----LGLIFIGVVPVIRKPGK 240
Db 451 ANKSLGQFNLG--DIAPAPRG 469

RESULT 11
ID DNAK PORGI STANDARD; PRT; 640 AA.
AC Q9ZAD3;
DT 16-OCT-2001 (Rel. 40, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DE 15-MAR-2004 (Rel. 43, Last annotation update)
DE Chapterone protein dnaK (Heat shock protein 70) (Heat shock 70 kDa protein) (HSP70).
GN DNAK OR PG1208.
OS Porphyromonas gingivalis (Bacteroides gingivalis).
OC Bacteria; Bacteroidetes; Bacteroides (Class); Bacteroidales;
OC Porphyromonadaceae; Porphyromonas.
OX NCBI_TaxID=837;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 33277;
RX MEDLINE=99198959; PubMed=10100860;
RA Yoshida A., Nakano Y., Yamashita Y., Oho T., Shibata Y., Ohishi M., Koga T.;
RT "A novel dnaK operon from Porphyromonas gingivalis.";
RL FEBS Lett. 446:287-291(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=W83;
RX MEDLINE=22829867; PubMed=12949112;
RA Nelson K.E., Fleischmann R.D., DeBoy R.T., Paulsen I.T., Fouts D.E., Eisen J.A., Dougherty S.C., Dodson R.J., Durkin A.S., Gwinn M., Haft D.H., Kolonay J.F., Nelson W.C., Mason J.L., Tallon L., Gray J., Granger D., Tettelin H., Dong H., Galvin J.L., Duncan M.J., Dewirst F.E., Fraser C.M.;
RT "Complete genome sequence of the oral pathogenic bacterium Porphyromonas gingivalis strain W83.";
RL J. Bacteriol. 185:5591-5601(2003).
CC -1- FUNCTION: Acts as a chaperone (By similarity).
CC -1- INDUCTION: By stress conditions e.g. heat shock (By similarity).
CC -1- SIMILARITY: Belongs to the heat shock protein 70 family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AB015879; BAA35087.1; -
DR EMBL; AE017176; AAQ66298.1; -
DR HSPF; P04475; 1D64.
DR TIGR; PG1208; -.
DR HAMAP; MF_00332; -.
DR InterPro; IPR001023; Hsp70.
DR Pfam; PF00012; HSP70; 1.
DR PRINTS; PR00301; HEATSHOCK70.
DR ProDom; PD000089; Hsp70; 1.
DR PROSITE; PS00297; HSP70_1; 1.
DR PROSITE; PS00329; HSP70_2; 1.
DR PROSITE; PS01036; HSP70_3; 1.
DR Chapterone; ATP-binding; Heat shock; Phosphorylation; Complete proteome.
KW MOD_RES 197
FT 197 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).

```

```

FT CONFLICT 169 169 E -> D (IN REF. 1).
FT CONFLICT 233 293 T -> R (IN REF. 1).
FT CONFLICT 235 295 S -> R (IN REF. 1).
FT CONFLICT 590 590 S -> T (IN REF. 1).
FT CONFLICT 600 600 A -> V (IN REF. 1).
FT CONFLICT 621 621 A -> V (IN REF. 1).
SQ SEQUENCE 640 AA; 68974 MW; 07CB789EFFFFEBC CRC64;

Query Match
Best Local Similarity 22.6%; Pred. No. 10;
Matches 54; Conservative 45; Mismatches 78; Indels 62; Gaps 13;

QY 6 VRADELVELQGLAESREQAKLIMAGKVTLTNNSTT-IPLRLKPGHKYVELSICSLIGV 64
Db 246 LQQPMA-MQRLKEAEKA-----KIELSTSTSTEINLPYMPVNGIPKHLVWIL---294
QY 65 ERFVSRGAYKLLTALDFFKID--VKSCIC-----LDAGASTGGFTDCLLOHGASKVYAI 116
Db 295 ---TRAKFEQLA-----DRLIQACVAPCETALKDAGMSRGDIDEVLVGSSTRIPAI 343
QY 117 DVGKQLHEKLYTNEQVINIEGVNLRASKOLIPBEV-----DILTIDVSFISLTL 167
Db 344 Q-----EIVEKIFG-----KAPSKGNPDEVAVGAATQGGVLTGEVKDVLID 387
QY 168 ILPSCIRWLKASGIIIALIKPQFELYPDKIKKGVKETSLOVE-AVEKIHFCQSELGI 225
Db 388 VTPLSLGIETMGVWTRLIDANTYI---PTKSEIFTAVDNQPSVE--IHVLQGERSL 441

RESULT 12
ID DNAK NITEU STANDARD; PRT; 644 AA.
AC O06430;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Chapterone protein dnaK (Heat shock protein 70) (Heat shock 70 kDa protein) (HSP70).
GN DNAK OR NE1949.
OS Nitrosomonas europaea.
OC Bacteria; Proteobacteria; Betaproteobacteria; Nitrosomonadales;
OC Nitrosomonadaceae; Nitrosomonas.
OX NCBI_TaxID=915;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 19718 / IFO 14298;
RX MEDLINE=97288071; PubMed=9143112;
RA Iizumi T., Nakamura K.;
RT "Cloning, nucleotide sequence, and regulatory analysis of the Nitrosomonas europaea dnaK gene.";
RL Appl. Environ. Microbiol. 63:1777-1784(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 19718 / IFO 14298;
RX MEDLINE=22586410; PubMed=12700255;
RA Chain P., Lamerdin J.E., Larimer F.W., Regala W., Lao V., Land M., Hauser L., Hooper A.B., Klotz M.G., Norton J., Sayavedra-Soto D.A., Arciero D.M., Hommes N.G., Whittaker M.M., Arp D.J.;
RT "Complete genome sequence of the ammonia-oxidizing bacterium and obligate chemolithoautotroph Nitrosomonas europaea.";
RL J. Bacteriol. 185:2759-2773(2003).
CC -1- FUNCTION: Acts as a chaperone (By similarity).
CC -1- INDUCTION: By stress conditions e.g. heat shock (By similarity).
CC -1- SIMILARITY: Belongs to the heat shock protein 70 family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
CC -----

```

```
DR EMBL; AB018706; BAA33935.1; --
DR EMBL; BX321862; CAD85860.1; --
DR HSSP; P04475; LDG4. --
DR HAMAP; MF_00332; --; 1.
DR InterPro; IPR001023; Hsp70.
DR Pfam; PF00012; HSP70; 1.
DR PRINTS; PR00301; HEATSHOCK70.
DR ProDom; PD000089; HSP70; 1.
DR PROSITE; PS00297; HSP70.1; 1.
DR PROSITE; PS00329; HSP70.2; 1.
DR PROSITE; PS01036; HSP70.3; 1.
DR Chaperone; ATP-binding; Heat shock; Phosphorylation;
KW Complete proteome.
FT MOD_RES 199 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
SQ SEQUENCE 644 AA; 69694 MW; 8E8BF50CE3B5D71C CRC64;

Query Match 6.8%; Score 87; DB 1; Length 644;
Best Local Similarity 23.3%; Pred. No. 10;
Matches 62; Conservative 36; Mismatches 78; Indels 90; Gaps 14;

QY 10 ELVFLQGLAESRQAKRLIMAGKVTLNNTTIP-LRLEKPGHKY-----PLESICS 60
DB 255 DMIALQRLKDAKAKIELSSQTEVN-----LPYTADASGPKHLAVKITRAKLES-- 308
QY 61 LIGVERFVSRGAYKLTALDFPKIDVKSCICLDAGSTGFTDCLLQHGASKYVAIDVGK 120
DB 309 ---VEELIERTAGPCRTALK-----DAGLSVSDINDVILVGGQTRMPKV---- 349
QY 121 GOLHEKLYTNEQVNIETGNLRTASKDLIPEEV-----DILTIDVSFI 163
DB 350 -----QEKVKEIFGKEPR---KDVNPDEAVAIGAAIQGGVLKGDVKKDVLDDVTP 397
QY 164 SL---TLILPSCIRWLKASGIITLIKPFQELYPDKIKGVKWKTSQYEAIVEKIHFCQ 220
DB 398 SLGIETL-----GGWTKLIQKNTTI-PTKAQQ-VFSTADNDQNAV--TIHVLO 442
QY 221 SE-----LGLIFIGVVPVSIK 237
DB 443 GEREVASGNKSLGQFNLTIPSAPRG 468

RESULT 13
MY5A MOUSE STANDARD; PRT; 1853 AA.
AC Q99104;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Myosin Va (Myosin 5A) (Dilute myosin heavy chain, non-muscle).
GN MYO5A OR DILUTE.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
[1]
SEQUENCE FROM N.A.
STRAIN=C57BL/6; TISSUE=Brain;
MEDLINE=91141583; PubMed=1996138;
RA Mercer J.A., Seperack P.K., Strobel M.C., Copeland N.G.,
RA Jenkins N.A.;
RA "Novel myosin heavy chain encoded by murine dilute coat colour
RT locus.";
RL Nature 349:709-712(1991).
RN [2]
RP REVISIONS.
RA Mercer J.A., Seperack P.K., Strobel M.C., Copeland N.G.,
RA Jenkins N.A.;
RA Nature 352:547-547(1991).
RN [3]
RP INTERACTION WITH MYRIP.
RX MEDLINE=22302022; PubMed=12221080;
RA Fukuda M., Kuroda T.S.;
RA "Slac2-c (synaptotagmin-like protein homologue lacking C2 Domains-c),
RT
```

```
RT a novel linker protein that interacts with Rab27, Myosin Va/Viia, and
RT actin.";
RL J. Biol. Chem. 277:43096-43103(2002).
RN [4]
RX INTERACTION WITH MLPH
RX MEDLINE=21941127; PubMed=11887186;
RA Wu X.S., Rao K., Zhang H., Wang F., Sellers J.R., Matesic L.E.,
RA Copeland N.G., Jenkins N.A., Hammer J.A. III;
RT "Identification of an organelle receptor for myosin-Va.";
RL Nat. Cell Biol. 4:271-278(2002).
CC -!- FUNCTION: Processive actin-based motor that can move in large
CC steps approximating the 36-nm pseudo-repeat of the actin filament.
CC Involved in melanosome transport. May also be required for some
CC polarization process involved in dendrite formation.
CC -!- SUBUNIT: May be a homodimer, which associates with multiple
CC calmodulin or myosin light chains. Binds MLPH and MYRIP.
CC -!- TISSUE SPECIFICITY: Detected in melanocytes.
CC -!- SIMILARITY: Contains 1 myosin-like globular head domain.
CC -!- SIMILARITY: Contains 1 dilute domain.
CC -!- SIMILARITY: Contains 6 IQ domains.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X57377; CAA40651.1; --
DR PIR; A46761; A46761.
DR HSSP; P10587; 1BR2.
DR MGD; MGI:105976; Myo5a.
DR InterPro; IPR002710; DIL.
DR InterPro; IPR000048; IQ region.
DR InterPro; IPR001609; myosin_head.
DR Pfam; PF01843; DIL; 1.
DR Pfam; PF00612; IQ; 6.
DR Pfam; PF00063; myosin_head; 1.
DR PRINTS; PR00193; MYOSINHEAVY.
DR ProDom; PD003376; DIL; 1.
DR ProDom; PD000355; myosin_head; 1.
DR SMART; SM00015; IQ; 6.
DR SMART; SM00242; MYSC; 1.
DR PROSITE; PSS0096; IQ; 6.
DR Myosin; Repeat; ATP-binding; Calmodulin-binding; Actin-binding;
KW Coiled coil; Phosphorylation.
FT DOMAIN 1 765 MYOSIN HEAD-LIKE.
FT DOMAIN 766 788 IQ 1.
FT DOMAIN 789 813 IQ 2.
FT DOMAIN 814 836 IQ 3.
FT DOMAIN 837 861 IQ 4.
FT DOMAIN 862 884 IQ 5.
FT DOMAIN 885 913 IQ 6.
FT DOMAIN 914 1237 COILED COIL (POTENTIAL).
FT DOMAIN 1314 1443 COILED COIL (POTENTIAL).
FT DOMAIN 1685 1790 DILUTE.
FT NP_BIND 163 170 ATP (POTENTIAL).
FT DOMAIN 643 665 ACTIN-BINDING (POTENTIAL).
FT MOD_RES 1758 1758 PHOSPHORYLATION (POTENTIAL).
SQ SEQUENCE 1853 AA; 215594 MW; 503E93D48CA6B766 CRC64;

Query Match 6.8%; Score 87; DB 1; Length 1853;
Best Local Similarity 21.6%; Pred. No. 34;
Matches 61; Conservative 46; Mismatches 93; Indels 82; Gaps 15;

QY 2 AKHKVRADELVFLQGLAESRQAKRLIMAGKVTLN-----NSTTIPRLRKPGHKY 53
DB 302 AKEMAHTRQACTLLGISSEYQMGIFRLIAGLILHLCNVGFASRSDSCTIP-----PKHE- 355
QY 54 PLESICSLIGVERFVSRGAY-----KLLFALDFPKIDVKSCICLDAGSTGGFTD 103
DB 356 PLTIFFCDLMGVD-----YEEMCHWLCHKRLATATETY---IKPISKLOATNA----RD 401
```

Qy	104	CLLQHGSKVY--AIDVGKGQLHE--KLYTNEQVINIGVNLRTASKOLIPBEVDILTID	159
Dd	402	ALAKHIYAKLPNTWIDHVNQALHSAAVKSHSFIVGLDIYGFE-----TFE	445
Qy	160	VSPFISLLIILPSCIRWLKASGIITIALIKIQPELYDPDKI-----KGVKBTSLOVEAVEK	214
Dd	446	INSFE-----QFCINYANEK-----LQQFNMFHVFKLEQEYMKEQIPWTLTDFDNDQP	494
Qy	215	IIFPCQSELGLFIFGWPSVIKPKGNQE-----YLIIYKK	250
Dd	495	CINLIIESKLGI--LDLDEECKMPKGTDDTWAQKLYNTHLNK	534
 RESULT 14 Y890_THEME			
ID	Y890_THEME	STANDARD;	PRT; 501 AA.
AC	Q9WZ22;		
DT	28-FEB-2003 (Rel. 41, Created)		
DT	28-FEB-2003 (Rel. 41, Last sequence update)		
DT	15-MAR-2004 (Rel. 43, Last annotation update)		
DE	Hypothetical zinc metalloprotease TM0890 (EC 3.4.24.-).		
GN	TM0890.		
OS	Thermotoga maritima.		
OC	Bacteria; Thermotogae; Thermotogales; Thermotogaceae; Thermotoga.		
NCBI_TaxID=2336;			
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=MGB8 / DSM 3109 / ATCC 43589;		
RX	MEDLINE=99287316; PubMed=10360571;		
RA	Nelson K.E., Clayton R.A., Gill S.R., Gwinn M.L., Dodson R.J.,		
RA	Haft D.H., Hickey E.K., Peterson J.D., Nelson W.C., Ketchum K.A.,		
RA	McDonald L., Utterback T.R., Malek J.A., Linher K.D., Garrett M.M.,		
RA	Stewart A.M., Cotton M.D., Pratt M.S., Phillips C.A., Richardson D.,		
RA	Heidelberg J., Sutton G.G., Fleischmann R.D., Eisen J.A., White O.,		
RA	Salzberg S.L., Smith H.O., Venter J.C., Fraser C.M.;		
RT	"Evidence for lateral gene transfer between Archaea and Bacteria from		
RL	genome sequence of Thermotoga maritima."		
RL	Nature 399:323-329(1999).		

```
--!- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane
-- (By similarity).
--!- SIMILARITY: Belongs to peptidase family M50B.
--!- SIMILARITY: Contains 1 PDZ/DHR domain.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to licenses@isb-sib.ch).
-----
EMBL; A8G01754; AAD35971.1; -.
PIR; C72321; C72321.
TIGR; TMO890; -.
InterPro; IPR001478; PDZ.
InterPro; IPR006025; Pept_M_Zn_BS.
InterPro; IPR008915; Peptidase_M50.
Pfam; PF00595; PDZ; 1.
Pfam; PF02163; Peptidase_M50; 1.
SMART; SM00228; PDZ; 1.
PROSITE; PS50106; PDZ; 1.
PROSITE; P500142; ZINC PROTEASE; 1.
KW Hypothetical protein; Hydrolase; Metalloprotease; Zinc; Transmembrane;
KW Inner membrane; Complete proteome.
FT METAL 17 17 ZINC (CATALYTIC) (POTENTIAL) .
FT ACT_SITE 18 18 POTENTIAL.
FT METAL 21 21 ZINC (CATALYTIC) (POTENTIAL) .
FT TRANSMEM 93 115 POTENTIAL.
FT TRANSMEM 401 420 POTENTIAL.
FT TRANSMEM 427 449 POTENTIAL.
FT TRANSMEM 474 496 POTENTIAL.
```

```

FT DOMAIN 96 180 PDZ.
SQ SEQUENCE 501 AA; 55875 MW; CE3E581117DC2A9A CRC64;

Query Match 6.8%; Score 86; DB 1; Length 501;
Best Local Similarity 25.68; Pred.No. 9.6; Mismatches 38; Indels 58; Gaps 13;
Matches 58; Conservative

QY 11 LVFIQGLAESREQAARLIMAGKVTLTNNSTTIPRLLEKPGHKYPLESICSL-IGVERFVS 69
Db 312 IVSLQG--ENIEWRWGLSGSVRVVVKRGDSTIEKNVEASFLKNILTFDLENGVPR--- 366
QY 70 RGAYKLLTALDFPKIDVKS--ICLDAGASTGFTCLLOHGASKYVAIDVGKGLHEKL 127
Db 367 ---YKPKNPLEANVLSYKACNYLLITASSL-----KNFRNVQTQI----- 406
QY 128 YTNEQVINIEGVNLRTASK--DLIPEVDILITDVSFISLTILPSCIRWLKASGIIIA 184
Db 407 ---VGWVGLAGV-ISAASKTGLAEVLTVAVITISLGVNL-LPLFA---LDGGRIFIS 457
QY 185 LIRPQPELYPDKIKGVKETSQYEAVERKIIHFQCSGELGIIFGV 231
Db 458 LVE-----MITRKLNPQ-VENIHF-----LGFILMIL 486

RESULT 15
MYSRAT ID MYSRAT STANDARD; PRT; 1828 AA.
AC Q9QYF3;
DT 16-OCT-2001 (Rel. 40, Created)
DT 18-OCT-2001 (Rel. 40, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Dilute Myosin Va (Myosin 5A) (Dilute myosin heavy chain, non-muscle).
GN MYO5A.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=101116;
RN [1]
RP MEDLINE=20380823; PubMed=10920234;
RX Futaki S., Takagishi Y., Hayashi Y., Ohmori S., Kanou Y., Inouye M.,
RA Oda S., Seo H., Iwakawa Y., Murata Y.;
RT "Identification of a novel myosin-Va mutation in an ataxic mutant
RT rat, dilute-opisthotonus."
RL Mamm. Genome 11:649-655(2000).
CC -/- FUNCTION: Processive actin-based motor that can move in large
CC steps approximating the 36-nm pseudo-repeat of the actin filament.
CC Involved in melanosome transport. May also be required for some
CC polarization process involved in dendrite formation (By
CC similarity).
CC -/- SUBUNIT: May be a homodimer, which associates with multiple
CC calmodulin or myosin light chains. Binds MLPH and MYRIP (By
CC similarity).
CC -/- DISEASE: Defects in MYO5A are a cause of Dilute-opisthotonus
CC (dop). Dop rats have diluted coat color and are occasionally
CC associated with severe neurological disorders.
CC -/- SIMILARITY: Contains 1 myosin-like globular head domain.
CC -/- SIMILARITY: Contains 6 IQ domains.
CC -/- SIMILARITY: Contains 1 dilute domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AB035736; BAA88350.1; -.
CC HSP; P10587; 1BR2.
CC InterPro; IPR002710; DIL.
CC InterPro; IPR000048; IQ region.
CC InterPro; IPR001609; myosin head.

```



```
DR PROSITE; PS50889; S4; 1.
KW Ribosomal protein; RNA-binding; rRNA-binding; Complete proteome.
FT DOMAIN 92 S4 RNA-BINDING.
SQ SEQUENCE 200 AA; 23040 MW; D64B3F3E497CFFC8 CRC64;

Query Match
Best Local Similarity 6.7%; Score 85.5; DB 1; Length 200;
Matches 29; Conservative 10; Mismatches 46; Indels 11; Gaps 2;

QY 7 RADELVFLQGLAESRQAKRLIMAGKVTLTNNSTTIPRLKPKGHKYPLESICSLIGVER 66
Db 93 RLDNLVIRLGLATRLRQAKQLVNHGVTVDGGVDPISVAIKPGQVIGLREKSNLDVVK 152

QY 67 FVSRGAYKLITLALDFPKIDVKSCICLDAGASTGGT 102
Db 153 -----EALNSF---VPEVTSFDADKMGESYT 177

RESULT 18
ID DNAK_DEIPR STANDARD; PRT; 618 AA.
AC P94695;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DE DnaK protein (HSP70) (Fragment).
DE protein (HSP70) (Heat shock protein 70) (Heat shock 70 kDa
GN DNAK.
OS Deinococcus proteolyticus.
OC Bacteria; Deinococcus-Thermus; Deinococci; Deinococcales;
OC Deinococcaceae; Deinococcus.
OX NCBI_TaxID=55148;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=ATCC 35074;
RX MEDLINE=97144518; PubMed=8990285;
RA Gupta R.S., Bustard K., Falah M., Singh D.;
RT "Sequencing of heat shock protein 70 (DnaK) homologs from Deinococcus
RT proteolyticus and Thermophilic proteobium roseum and their integration in a
RT protein-based phylogeny of prokaryotes.";
RL J. Bacteriol. 179:345-357(1997).
CC -1- FUNCTION: Acts as a chaperone (By similarity).
CC -1- INDUCTION: By stress conditions e.g. heat shock (By similarity).
CC -1- SIMILARITY: Belongs to the heat shock protein 70 family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC ENBL; U80215; AAB41739.1; -.
DR HSSP; P04475; IDG4.
DR HAMAP; MF_00332; -.
DR InterPro; IPR001023; Hsp70.
DR Pfam; PF00012; HSP70; 1.
DR PRINTS; PR00301; HEATSHOCK70.
DR PRODOM; PD000089; Hsp70; 1.
DR PROSITE; PS00297; HSP70_1; PARTIAL.
DR PROSITE; PS00329; HSP70_2; 1.
DR PROSITE; PS01036; HSP70_3; 1.
KW Chaperone; ATP-binding; Heat shock; Phosphorylation.
FT NON_TER 1 1
FT MOD_RES 184 184 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
SQ SEQUENCE 618 AA; 67206 MW; B21142B465E678F3 CRC64;

Query Match
Best Local Similarity 6.7%; Score 85.5; DB 1; Length 618;
Matches 47; Conservative 27; Mismatches 55; Indels 39; Gaps 9;

QY 14 LQGLAESRQAKRLIMAGKVTLTNNSTT---IPRLKPKGHKYPLESICSLIGVERFVSR 70
```

```
Db 240 LQRLTEAARKA-----KIELNSASETSIPLFTTFDPETRTPLH-----LERTLSR 285
QY 71 GAYKLLTALDFFKIDVKSCI---CLDAGASTGGTDCILQHGASKVYVAD-----VGKG 121
Db 286 AKFEELTA-DLLK-RVRQPVEQMRDAGVSSDLNEVLVGGSTRIPAVKRIYKDLTKG- 342
QY 122 QLHEKLYTNEQVINTIEGVNLTASKDLIPEE-----VDILTIDVSFISL 165
Db 343 -----EPNESVNEDEAVGLGAAVQAGIIQGDNSLGDIVLVDVTPLTL 384

RESULT 19
ID RS4_BACHD STANDARD; PRT; 200 AA.
AC Q9K728;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DE 30S ribosomal protein S4.
DE RPSD OR BH3209.
GN Bacillus halodurans.
OS Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=86665;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=C-125 / JCM 9153;
RX MEDLINE=20512582; PubMed=11058132;
RA Takami H., Nakasone K., Takaki Y., Maeno G., Sasaki R., Masui N.,
RA Fuji F., Hirama C., Nakamura Y., Ogasawara N., Kuhara S.,
RA Horikoshi K.;
RT "Complete genome sequence of the alkaliphilic bacterium Bacillus
RT halodurans and genomic sequence comparison with Bacillus subtilis.";
RL Nucleic Acids Res. 28:4317-4331(2000).
CC -1- FUNCTION: One of the primary rRNA binding proteins, it binds
CC directly to 16S rRNA where it nucleates assembly of the body of
CC the 30S subunit (By similarity).
CC -1- FUNCTION: With S5 and S12 plays an important role in translational
CC accuracy (By similarity).
CC -1- SUBUNIT: Part of the 30S ribosomal subunit. Contacts protein S5.
CC The interaction surface between S4 and S5 is involved in control
CC of translational fidelity (By similarity).
CC -1- SIMILARITY: Contains 1 S4 RNA-binding domain.
CC -1- SIMILARITY: Belongs to the S4P family of ribosomal proteins.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC ENBL; AP001518; BAB06928.1; -.
DR PIR; A84051; A84051.
DR HSSP; P81288; 1C05.
DR HAMAP; MF_01306; -.
DR InterPro; IPR001912; Ribosomal_S4.
DR InterPro; IPR005709; Ribosomal_S4_b/o.
DR InterPro; IPR002942; S4.
DR Pfam; PF00163; Ribosomal_S4; 1.
DR Pfam; PF01479; S4; 1.
DR SMART; SM00363; S4; 1.
DR TIGRFAMs; TIGR01017; rpsd bact; 1.
DR PROSITE; PS00632; RIBOSOMAL_S4; 1.
DR PROSITE; PS50889; S4; 1.
KW Ribosomal protein; RNA-binding; Complete proteome.
FT DOMAIN 92 152 S4 RNA-BINDING.
SQ SEQUENCE 200 AA; 23034 MW; F30CAB82787CD00FE CRC64;

Query Match
Best Local Similarity 6.6%; Score 84.5; DB 1; Length 200;
Matches 30; Conservative 14; Mismatches 38; Indels 17; Gaps 4;
```



```
ANM4 HUMAN
ID ANM4 HUMAN STANDARD; PRT; 334 AA.
AC Q9NR22;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Protein arginine N-methyltransferase 4 (EC 2.1.1.-).
GN HRMT1L3.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RC Lorenz B., Strom T.M.;
RT "Transcripts in human map region 12p13.3";
RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: Probably methylates the guanidino nitrogens of arginyl
CC residues in some proteins (By similarity).
CC -!- SIMILARITY: Belongs to the protein arginine N-methyltransferase
CC family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AF263539; AAF91390.1; -.
CC Genew; HGNC:5188; HRMT1L3.
CC InterPro; IPR000051; SAM_bind.
CC Transferrase; Methyltransferase.
SQ SEQUENCE 334 AA; 38835 MW; C3BC4373159B224A CRC64;

Query Match 6.6%; Score 84; DB 1; Length 334;
Best Local Similarity 22.8%; Pred. No. 8.9;
Matches 37; Conservative 29; Mismatches 72; Indels 24; Gaps 6;

QY 87 KSCICIDAGASTCGFTDCLLOHGASKVYVIDVCK-GQLHEKLYTNEQVINIEGV-NLRTA 144
DB 51 KDKVLDVSGTGTILSMFAKAGAKVFGIECSSISDYSEKIIKANELDNIIIFKGVKE 110
QY 145 SKDLIPEEVNLDIVSFSLTILPSCI-----RWLKASGIIIALIKPQFELYPPK 196
DB 111 EVELPEKVDIIISE--WMGYCLFYESMLNTVIFARDKWLKPGSL-----MEPDR 158
QY 197 IKGVVVKETSLQYEAIVEKIHFQSELGLIFIGVWPSVIKGP 238
DB 159 AALYVVAIEDRQYKDFK-IHWENVYGFDMTCIRDVAMKEP 198

RESULT 23
RS4 STRPN STANDARD; PRT; 203 AA.
AC Q97T69;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE 30S ribosomal protein S4.
GN RPSD OR SP0085 OR SP0078.
OS Streptococcus pneumoniae, and
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1313, 171101;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC BAA-334 / TIGR4;
RC MEDLINE=21357209; PubMed=11463916;
RX Tettelin H., Nelson K.E., Paulsen I.T., Eisen J.A., Read T.D.,
RA
```



```
RESULT 24
RS4_STRPY STANDARD; PRT; 203 AA.
AC O95XJ4;
DT 28-FEB-2003 (Rel. 41, Last Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE 30S ribosomal protein S4.
GN RPSD OR SPY2178 OR SPYM3_1833 OR SPS1829 OR SPYM18_2215.
OS Streptococcus pyogenes, (serotype M3), and
OS Streptococcus pyogenes.
OC Streptococcus pyogenes (serotype M18).
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1314, 198466, 186103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SF370 / ATCC 700294 / Serotype M1;
RX MEDLINE=21192684; PubMed=11296296;
RA Ferretti J.J., McShan W.M., Ajdic D.J., Savic D.J., Savic G., Lyon K.,
RA Primeaux C., Sezate S., Suvorov A.N., Kenton S., Lai H.S., Lin S.P.,
RA Qian Y., Jia H.G., Najjar F.Z., Ren Q., Zhu H., Song L., White J.,
RA Yuan X., Clifton S.W., Roe B.A., McLaughlin R.;
RA "Complete genome sequence of an M1 strain of Streptococcus pyogenes.";
RA Proc. Natl. Acad. Sci. U.S.A. 98:4658-4663(2001).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=MGAS315 / Serotype M3;
RX MEDLINE=22133808; PubMed=12122206;
RA Beres S.B., Sylva G.L., Barbican K.D., Lei B., Hoff J.S.,
RA Mammarella N.D., Liu M.-Y., Smoot J.C., Porcella S.F., Parkins L.D.,
RA Campbell D.S., Smith T.M., McCormick J.K., Leung D.Y.M.,
RA Schlievert P.M., Musser J.M.;
RA "Genome sequence of a serotype M3 strain of group A Streptococcus:
RA phage-encoded toxins, the high-virulence phenotype, and clone
RA emergence.";
RA Proc. Natl. Acad. Sci. U.S.A. 99:10078-10083(2002).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=SSI-1 / Serotype M3;
RX MEDLINE=22683278; PubMed=12799345;
RA Nakagawa I., Kurokawa K., Yamashita A., Nakata M., Tomiyasu Y.,
RA Okahashi N., Kawabata S., Yamazaki K., Shiba T., Yasunaga T.,
RA Hayashi H., Hattori M., Hamada S.;
RA "Genome sequence of an M3 strain of Streptococcus pyogenes reveals a
RA large-scale genomic rearrangement in invasive strains and new insights
RA into phage evolution.";
RA Genome Res. 13:1042-1055(2003).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=MGAS8232 / Serotype M18;
RX MEDLINE=21927593; PubMed=11917108;
RA Smoot J.C., Barbican K.D., Van Gompel J.J., Smoot L.M., Chaussee M.S.,
RA Sylva G.L., Sturdevant D.E., Ricklefs S.M., Porcella S.F.,
RA Parkins L.D., Beres S.B., Campbell D.S., Smith T.M., Zhang Q.,
RA Kapur V., Daly J.A., Veasy L.G., Musser J.M.;
RA "Genome sequence and comparative microarray analysis of serotype M18
RA group A Streptococcus strains associated with acute rheumatic fever
RA outbreaks.";
RA Proc. Natl. Acad. Sci. U.S.A. 99:4668-4673(2002).
CC -!- FUNCTION: One of the primary rRNA binding proteins, it binds
CC directly to 16S rRNA where it nucleates assembly of the body of
CC the 30S subunit (By similarity).
CC -!- FUNCTION: With S5 and S12 plays an important role in translational
CC accuracy (By similarity).
CC -!- SUBUNIT: Part of the 30S ribosomal subunit. Contacts protein S5.
CC The interaction surface between S4 and S5 is involved in control
CC of translational fidelity (By similarity).
CC -!- SIMILARITY: Contains 1 S4 RNA-binding domain.
CC -!- SIMILARITY: Belongs to the S4P family of ribosomal proteins.
CC

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).

DR EMBL; AE006635; AAK34809.1; -
DR EMBL; AE014172; AAM80440.1; -
DR EMBL; AP005146; BAC64924.1; -
DR EMBL; AB010121; AAL98649.1; -
DR HSSP; P81288; 1C05.
DR HAMAP; MF_01306; -; 1.
DR InterPro; IPR001912; Ribosomal_S4.
DR InterPro; IPR005709; Ribosomal_S4_b/o.
DR InterPro; IPR002942; S4.
DR Pfam; PF00163; Ribosomal_S4; 1.
DR Pfam; PF01479; S4; 1.
DR TIGRFAMs; TIGR01017; rpsD_bact; 1.
DR PROSITE; PS00632; RIBOSOMAL_S4; 1.
DR PROSITE; PS00889; S4; 1.
KW Ribosomal protein; rRNA-binding; rRNA-binding; Complete proteome.
FT DOMAIN 93 156 S4 RNA-BINDING.
SQ SEQUENCE 203 AA; 23122 MW; BE4E99E34D7D791D CRC64;

Query Match 6.5%; Score 83; DB 1; Length 203;
Best Local Similarity 32.2%; Pred. No. 6.2; Mismatches 12; Indels 14; Gaps 3;
Matches 28; Conservative 12;

QY 7 RADELVFLQGLAESREQKRLIMAGKVTLTNNSTTLRLLEKPGH-----KYP--L 55
DB 94 RLDNVVRLGLATTRQARQVNHGHIIVDGKRVDFISYVDPGQVLSVREKSMKVPAIL 153
QY 56 ESICSLIGLGVFVSRGAYKL---LTAL 79
DB 154 EAVEATLGRPAFVSFDAEKLEGLSUTRL 180

RESULT 25
KPR4_YEAST STANDARD; PRT; 355 AA.
AC P38063;
DT 01-OCT-1994 (Rel. 30, Created)
DT 01-OCT-1994 (Rel. 30, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Probable ribose-phosphate pyrophosphokinase 4 (EC 2.7.6.1)
DE (Phosphoribosyl pyrophosphate synthetase 4).
GN PRPS4 OR PRS4 OR YBL068W OR YBL0619.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C;
RA Contreras R., Fiers W., Logghe M., Molemans F.;
RA Submitted (AUG-1994) to the EMBL/GenBank/DBJ databases.
CC -!- CATALYTIC ACTIVITY: ATP + D-ribose 5-phosphate = AMP + 5-phospho-
CC alpha-D-ribose 1-diphosphate.
CC -!- SIMILARITY: Belongs to the ribose-phosphate pyrophosphokinase
CC family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; Z35829; CAA84888.1; -
CC PIR; S45804; S45804.
```

```
DR HSP, P14193; LDKU.
DR GERMOnline; 138503; -.
DR SGD; S0000164; PR94.
DR InterPro; IPR000842; PRPP_synthetase.
DR InterPro; IPR000836; PRTransferrase.
DR InterPro; IPR005946; RibP Ppkin.
DR Pfam; PF00156; Ribosyltran; 1.
DR TIGRFAMS; TIGR01251; ribP Ppkin; 1.
DR PROSITE; PS00114; PRPP SYNTHETASE; 1.
KW Nucleotide biosynthesis; Transferase; Kinase; Magnesium;
KW Multigene family.
FT METAL 168 168 MAGNESIUM (POTENTIAL).
FT METAL 170 170 MAGNESIUM (POTENTIAL).
FT METAL 179 179 MAGNESIUM (POTENTIAL).
FT METAL 183 183 MAGNESIUM (POTENTIAL).
SQ SEQUENCE 355 AA; 33059 MW; 41EBD027E5399B46 CRC64;

Query Match 6.5%; Score 83; DB 1; Length 355;
Best Local Similarity 31.7%; Pred. No. 12;
Matches 33; Conservative 16; Mismatches 43; Indels 12; Gaps 5;

QY 60 SLIGVERFVSRGAYKLLTALDFKIDVKSCICLDAGASTGG---FTDCLQHGASKVYA 115
Db 230 ALIHGRQKANEVSKWLVGD---VTNKSCLLDVDDADTCGLTKACDILMEHGAKEVIA 286

QY 116 I---DVKGQGLHEKLYTNEQVINIEGVNLRTASKDL-IPDEVPI 155
Db 287 IVTHGIFSGSAREKL-RNSRLSRIVCTNTPVDLDLPADQIDI 329

RESULT 26
GFR2 HUMAN
ID_GFR2_HUMAN STANDARD; PRT; 464 AA.
AC 000451; O15328;
DT 01-NOV-1997 (Rel. 35, Created)
DT 10-NOV-1997 (Rel. 35, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE GDNF family receptor (GFR-alpha 2) (Neurturin
DE receptor alpha) (NTRN-alpha) (NTRN-alpha) (TGF-beta related
DE neurotrophic factor receptor 2) (GDNF receptor beta) (RET
DE ligand 2).
DE GFR2 OR GDNFR OR TRNR2 OR RETL2.
GN Homo sapiens (human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).
RX MEDLINE=97325791; PubMed=9182803;
RA Baloh R.H., Tansey M.G., Golden J.P., Creedon D.J.,
RA Heuckeroth R.O., Keck C.L., Zimonjic D.B., Popescu N.C.,
RA Johnson E.M. Jr., Milbrandt J.;
RT "Trnr2, a novel receptor that mediates neurturin and GDNF signaling
RT through Ret.";
RL Neuron 18:793-802(1997).
[2]
SEQUENCE FROM N.A. (ISOFORM 1).
RC TISSUE=Fetal brain;
RX MEDLINE=97402208; PubMed=9259272;
RA Suvanto P., Warciovaara K., Lindahl M., Arumae U., Moshnyakov M.,
RA Horelli-Kuitunen N., Aliraksinen M.S., Palotie A., Sariola H.,
RA Saarma M.;
RT "Cloning, mRNA distribution and chromosomal localisation of the gene
RT for glial cell line-derived neurotrophic factor receptor beta, a
RT homologue to GDNFR-alpha.";
RL Hum. Mol. Genet. 6:1267-1273(1997).
[3]
SEQUENCE FROM N.A. (ISOFORM 1).
RC TISSUE=Liver;
RX MEDLINE=97322356; PubMed=9177201;
RA Sanicola M., Hession C.A., Worley D.S., Carmillo P., Ehrenfels C.,
RA Walus L., Robinson S., Jaworski G., Wei H., Tizard R., Whitty A.,
RA Pepinsky R.B., Cate R.L.;
```

```
RT "Glial cell line-derived neurotrophic factor-dependent RET activation
RT can be mediated by two different cell-surface accessory proteins.";
RL Proc. Natl. Acad. Sci. U.S.A. 94:6238-6243(1997).
CC -!- FUNCTION: Receptor for neurturin. Mediates the NRTN-induced
CC autophosphorylation and activation of the RET receptor. Also able
CC to mediate GDNF signaling through the RET tyrosine kinase
CC receptor.
CC -!- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor (By
CC similarity).
CC -!- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=2;
CC Name=1; Synonyms=Long;
CC IsoId=O00451-1; Sequence=Displayed;
CC Name=2; Synonyms=Short;
CC IsoId=O00451-2; Sequence=VSP_001661;
CC -!- TISSUE SPECIFICITY: Isoform 1 is found in both brain and placenta.
CC -!- SIMILARITY: Belongs to the GDNFR family.
CC
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AF002700; AAC52036.1; -.
CC EMBL; U93703; AAB61922.1; -.
CC EMBL; U97145; AAC51647.1; -.
CC Genew; HGNC:4244; GFR2.
CC MIM; 601956; -.
CC GO; GO:0016167; F:glial cell line-derived neurotrophic factor. .; TAS.
CC GO; GO:0015025; F:GPI-anchored membrane-bound receptor; TAS.
CC GO; GO:0007169; P:transmembrane receptor protein tyrosine kin. .; TAS.
CC InterPro; IPR003438; GDNF_receptor.
CC Pfam; PF02351; GDNF; 1.
CC PRINTS; PR0316; GDNFRECEPTOR.
KW Receptor; Glycoprotein; GPI-anchor; Membrane; Signal;
KW Alternative splicing; Lipoprotein.
FT SIGNAL 1 21 POTENTIAL.
FT CHAIN 22 444 GDNF FAMILY RECEPTOR ALPHA 2.
FT PROPEP 445 464 REMOVED IN MATURE FORM (POTENTIAL).
FT CARBOHYD 52 52 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 357 357 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 413 413 N-LINKED (GLCNAC. .) (POTENTIAL).
FT LIPID 444 444 GPI-anchor amidated serine (Potential).
FT VARSPLOC 14 146 Missing (in isoform 2).
FT /FTId=VSP_001661.
FT CONFLICT 6 6 V -> A (IN REF. 2).
FT CONFLICT 462 462 Q -> L (IN REF. 3).
SQ SEQUENCE 464 AA; 51558 MW; 8BC604D9530FF21F CRC64;

Query Match 6.5%; Score 83; DB 1; Length 464;
Best Local Similarity 22.3%; Pred. No. 16;
Matches 60; Conservative 23; Mismatches 94; Indels 92; Gaps 13;

QY 6 VRADELIVFLOGLAESREQAARLIMAGK---VTITNSTTIPDLRL--EKP-----GHX 52
Db 41 VRANELCAEBSNCSSRYRTLRQCLAGDRNRLTLANKECQAALVQLQESPLYDCRCRGMK 100
QY 53 YPLESI-----CSLIGVERFVSRGAYKLLTA--LDVFEK-----IDVKSCL 90
Db 101 KEJQCQIQIYWSIHGLGTGEFEFVSPVPTSLRSLDFIRLASFSTGADPVVSAKSNH 160
QY 91 CLDAGASTGGFTDCLQHGASKVYAIDVGKQLHEKLYTNEQVINIEGVNLRTASK--- 146
Db 161 CLDAKA-----CNLDNCKLRSYIS-----ICNREISPTERCNRKCHKALRQ 206
QY 147 --DLIPEEVDILITDVSFISL-----TILPSCIRWLKASGIIILIKPQLPLYP 194
Db 207 FFDVRPSE---YTYRMLFSCQDQACAEERRRQITLPSC-----SYE 244
QY 195 DKIKGVVVKETSLQYEAVERKIIHFQCSQL 223
```


Db 348 -----QKQVADFFG-----KTRKDVNPDEAVMAIGAAGVQGVLSGVSVDVLLDVT 394
QY 163 ISLTILPSCIRWLKASGIIIALIKPQFELYPDKIKKGVVKETSLQYEAIVEKIHFQCE 222
Db 395 LSLGI-----ETWGVVTTILIEKNVTI---PTKKSQVFSTAEDNQSA-VIIHVLQGE 442

RESULT 29
ID BPAL HUMAN STANDARD; PRT: 3214 AA.
AC Q03001; Q12825; Q13266; Q13267; Q13775; Q96J76; Q96QT5; Q9UGD7;
AC Q9UGD8;
DT 01-OCT-1996 (Rel. 34, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Bullous pemphigoid antigen 1 isoforms 1/2/3/4/5/8 (230 kDa bullous
DE pemphigoid antigen) (BPA) (Hemidesmosomal plaque protein) (Dystonia
DE musculorum protein) (Fragment).
GN BPAG1 OR DMH OR DT OR KIAA0728.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORM 3).
RC TISSUE=Keratinocytes;
RX MEDLINE=92011493; PubMed=1717441;
RA Sawamura D., Li K., Chu M.-L., Uitto J.;
RA "Human bullous pemphigoid antigen (BPAG1). Amino acid sequences
RT deduced from cloned cDNAs predict biologically important peptide
RT segments and protein domains.";
RL J. Biol. Chem. 266:17784-17790(1991).
RN [2]
RP SEQUENCE FROM N.A. (ISOFORM 3).
RC TISSUE=Keratinocytes;
RX MEDLINE=93346806; PubMed=8345227;
RA Elgart G.W., Stanley J.R.;
RA "Cloning of the 5' mRNA for the 230-kD bullous pemphigoid antigen by
RT rapid amplification of cDNA ends.";
RL J. Invest. Dermatol. 101:244-246(1993).
RN [3]
RP SEQUENCE OF 1-645 FROM N.A. (ISOFORMS 1 AND 2).
RC TISSUE=Fetal brain, and Retina;
RX MEDLINE=96121394; PubMed=8575775;
RA Brown A., Dalpe G., Mathieu M., Kothary R.;
RA "Cloning and characterization of the neural isoforms of human
RT dystonin.";
RL Genomics 29:777-780(1995).
RN [4]
RP SEQUENCE OF 321-3214 FROM N.A. (ISOFORMS 1 AND 4).
RA Laird G.;
RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
RN [5]
RP SEQUENCE OF 392-492 FROM N.A.
RC TISSUE=Pineal gland;
RA Geerts D., Sonnenberg A.;
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
RN [6]
RP SEQUENCE OF 1082-3214 FROM N.A. (ISOFORM 3).
RC TISSUE=Keratinocytes;
RX MEDLINE=91286285; PubMed=1712022;
RA Tanaka T., Farry D.A.D., Klaus-Kovtun V., Steinert P.M.,
RA Stanley J.R.;
RT "Comparison of molecularly cloned bullous pemphigoid antigen to
RT desmoplakin I confirms that they define a new family of cell adhesion
RT junction plaque proteins.";
RL J. Biol. Chem. 266:12555-12559(1991).
RN [7]
RP SEQUENCE OF 2160-2767 FROM N.A.
RX MEDLINE=91216368; PubMed=2090522;
RA Owaribe K., Kastenbeck J., Stumpp S., Magin T.M., Krieg T.,
RA Diaz L.A., Franke W.W.;
RT "The hemidesmosomal plaque. I. Characterization of a major

RT constituent protein as a differentiation marker for certain forms of
RT epithelia.";
RL Differentiation 45:207-220(1990).
RN [8]
RP SEQUENCE OF 2287-3214 FROM N.A. (ISOFORM 5).
RC TISSUE=Keratinocytes;
RX MEDLINE=89067122; PubMed=2461961;
RA Stanley J.R., Tanaka T., Mueller S., Klaus-Kovtun V., Roop D.;
RA "Isolation of complementary DNA for bullous pemphigoid antigen by use
RT of patients' autoantibodies.";
RL J. Clin. Invest. 82:1864-1870(1988).
RN [9]
RP SEQUENCE OF 2462-3214 FROM N.A. (ISOFORM 8).
RX MEDLINE=94280413; PubMed=8010969;
RA Hopkinson S.B., Jones J.C.;
RT "Identification of a second protein product of the gene encoding a
RT human epidermal autoantigen.";
RL Biochem. J. 300:851-857(1994).
RN [10]
RP DOMAINS.
RX MEDLINE=96199235; PubMed=8621649;
RA Tang H.-Y., Chaffotte A.-F., Thacher S.M.;
RA "Structural analysis of the predicted coiled-coil rod domain of the
RT cytoplasmic bullous pemphigoid antigen (BPAG1). Empirical
RT localization of the N-terminal globular domain-rod boundary.";
RL J. Biol. Chem. 271:9716-9722(1996).
RN [11]
RP SUBCELLULAR LOCATION, TISSUE SPECIFICITY, AND ALTERNATIVE SPLICING.
RC TISSUE=Keratinocytes;
RX MEDLINE=21839111; PubMed=11751855;
RA Okumura M., Yamakawa H., Ohara O., Owaribe K.;
RA "Novel alternative splicings of BPAG1 (bullous pemphigoid antigen 1)
RT including the domain structure closely related to MACF (microtubule
RT actin cross-linking factor).";
RL J. Biol. Chem. 277:6682-6687(2002).
CC -I- FUNCTION: Cytoskeletal linker protein. Anchors keratin-containing
CC intermediate filaments to the inner plaque of hemidesmosomes. The
CC proteins may self-aggregate to form filaments or a two-dimensional
CC mesh.
CC -I- SUBUNIT: Homodimer.
CC -I- SUBCELLULAR LOCATION: Cytoplasmic.
CC -I- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=10;
CC Comment=Isoforms 1, 2, 5 and 8 are or may be fragments;
CC Name=1;
CC IsoId=Q03001-1; Sequence=Displayed;
CC Name=2;
CC IsoId=Q03001-2; Sequence=VSP_005053, VSP_005055;
CC Name=3; Synonyms=1e;
CC IsoId=Q03001-3; Sequence=VSP_005054, VSP_005056, VSP_005057,
CC VSP_005058, VSP_005059, VSP_005060,
CC VSP_005061;
CC Name=4;
CC IsoId=Q03001-4; Sequence=VSP_005054, VSP_005056;
CC Name=5;
CC IsoId=Q03001-5; Sequence=VSP_005064, VSP_005065;
CC Name=6; Synonyms=EA;
CC IsoId=Q04833-2; Sequence=External;
CC Name=7; Synonyms=EB;
CC IsoId=Q08WXK8-1; Sequence=External;
CC Name=8;
CC IsoId=Q03001-6; Sequence=VSP_005062, VSP_005063;
CC Name=9;
CC IsoId=Q04833-3; Sequence=External;
CC Name=10;
CC IsoId=Q04833-1; Sequence=External;
CC Note=No experimental confirmation available;
CC -I- TISSUE SPECIFICITY: Highly expressed in skeletal muscle and
CC cultured keratinocytes.
CC -I- DISEASE: BPAG1 is an autoantigen of bullous pemphigoid
CC [MIM:600086], an autoimmune subepithelial skin blistering disease.
CC -I- SIMILARITY: Belongs to the plakins or cytokerlin family.
CC -I- SIMILARITY: Contains 1 actin-binding domain.

```
CC -!- SIMILARITY: Contains 2 calponin-homology (CH) domains.
CC -!- SIMILARITY: Contains 1 SH3 domain.
CC -!- SIMILARITY: Contains 11 plectrin repeats.
CC -!- SIMILARITY: Contains 4 spectrin repeats.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M69225; -; NOT ANNOTATED_CDS.
CC EMBL; L11690; AAA52288.1; -
CC EMBL; U31850; AAC50243.1; -
CC EMBL; U31851; AAC50244.1; -
CC EMBL; AL096710; -; NOT ANNOTATED_CDS.
CC EMBL; AY032900; AAK63130.1; -
CC EMBL; AY032901; AAK63131.1; -
CC EMBL; M63618; AAA35606.1; -
CC EMBL; X58677; CAA41528.1; -
CC EMBL; M22942; AAA35538.1; -
CC EMBL; U04850; AAA57184.1; -
CC EMBL; U04850; AAA57185.1; -
CC PIR; I56317; A40937.
CC Genew; HGNC:1090; BFAG1.
CC MIM; 113810; -
CC MIM; 600888; -
CC GO; GO:0005604; C:basement membrane; TAS.
CC GO; GO:0005737; C:cytoplasm; IEP.
CC GO; GO:0005200; F:structural constituent of cytoskeleton; IEP.
CC GO; GO:0007010; P:cytoskeleton organization and biogenesis; TAS.
CC GO; GO:0045104; P:intermediate filament cytoskeleton organiza. . .; IEP.
CC InterPro; IPR001589; Actbind.actinin.
CC InterPro; IPR001715; Calponin-like.
CC InterPro; IPR001452; SH3.
CC InterPro; IPR002017; Spectrin.
CC Pfam; PF00307; CH; 2.
CC Pfam; PF00681; Plectrin; 5.
CC Pfam; PF00018; SH3; 1.
CC Pfam; PF00435; spectrin; 3.
CC SMART; SM00033; CH; 2.
CC SMART; SM00150; SPEG; 2.
CC PROSITE; PS00002; SH3; FALSE NEG.
CC PROSITE; PS00019; ACTININ_1; 1.
CC PROSITE; PS00020; ACTININ_2; FALSE NEG.
CC PROSITE; PS00021; CH; 2.
CC Antigen; Actin-binding; Coiled coil; Repeat; SH3 domain;
CC Structural protein; Cytoskeleton; Cell adhesion; Calcium;
CC Calcium-binding; Alternative splicing.
CC NON_TER 1 1
CC DOMAIN 1 1669
CC FT DOMAIN 1670 2441 GLOBULAR 1.
CC FT DOMAIN 2442 3214 CENTRAL FIBROUS ROD DOMAIN.
CC FT DOMAIN 215 439 GLOBULAR 2.
CC FT DOMAIN 219 322 ACTIN-BINDING.
CC FT DOMAIN 335 436 CH 1.
CC FT DOMAIN 335 436 CH 2.
CC FT REPEAT 774 851 SPECTRIN 1.
CC FT REPEAT 876 959 SPECTRIN 2.
CC FT REPEAT 984 1077 SPECTRIN 3.
CC FT DOMAIN 1091 1143 SH3.
Query Match 6.5%; Score 83; DB 1; Length 3214;
Best Local Similarity 24.4%; Pred. No. 1.4e+02;
Matches 53; Conservative 29; Mismatches 83; Indels 52; Gaps 9;
QY 31 GKVLITNNSTPIPLR-----LEKPGHKYPLESICSLIGVERFVSRYGAYKLLTALDFPKI 84
Db 2968 GLITITELADELLSLVLPKKDLHSPVAGYWLITASGERISLVKASRRNVDRITALR----- 3043
QY 85 DVKSCICILDAGASTGGTDCLLQHGASKVY-----AIDVGKQLHEKLJ----- 127
```

```
Db 3044 -----CLEARQVSTGGIIDL-----TGKRYRVAEALHRLGLVDEFGAQLRQCELVITGIG 3093
QY 128 --VTNEQVINTEGVLNLTASKDLIPERVDILTIDVFSILTLILPSCIRMLKAS-----G 180
Db 3094 HPITNKWMSVVEAVN-----ANIINKEMGRCUEFQYLTGLTIEPQVHSRLSIEALQVG 3148
QY 181 IIALIKLPQPELYPKDKIKGVVKET--SLOY-EAVEK 214
Db 3149 IIDVLIATKLKQDSYVNIICPTQTKRKLTKALEK 3185
RESULT 30
RS4_BACSU STANDARD; PRT; 199 AA.
AC P21466; Q45662; Q53282;
DT 01-MAY-1991 (Rel. 18, Created)
DT 01-AUG-1991 (Rel. 19, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE 30S ribosomal protein S4 (BS4).
GN RPSD OR BSU29660.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=91035248; PubMed=1699930;
RT Grundy F.J., Henkin T.M.;
RT "Cloning and analysis of the Bacillus subtilis rpsD gene, encoding
RT ribosomal protein S4."
RL J. Bacteriol. 172:6372-6379(1990).
[2]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=98048467; PubMed=9387221;
RA Lapidus A., Galleron N., Sorokin A., Ehrlich S.D.;
RA "Sequencing and functional annotation of the Bacillus subtilis genes
RA in the 200 kb rnbB-dnaB region."
RT Microbiology 143:3431-3441(1997).
[3]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=98044033; PubMed=93843377;
RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
RA Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,
RA Borriess R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,
RA Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,
RA Choi S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A.,
RA Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T.,
RA Entian K.D., Erington J., Fabret C., Ferrari E., Foulger D.,
RA Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
RA Ghim S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,
RA Giuseppe G., Guy B.J., Haga K., Haiech J., Harwood C.R., Henaut A.,
RA Hilbert H., Holsappel S., Hosono S., Hulio M.F., Itaya M., Jones L.,
RA Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,
RA Kobayashi Y., Koetter P., Koningsstein G., Krogh S., Kumano M.,
RA Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
RA Lee S.M., Levine A., Liu H., Masuda S., Maue C., Medigue C.,
RA Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,
RA Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,
RA Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,
RA Presecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
RA Rieger M., Rivolta C., Roche E., Roche B., Rose M., Sadaie Y.,
RA Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,
RA Sekiguchi J., Sekowska A., Seror S.J., Serror P., Shin B.S., Soldo B.,
RA Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,
RA Takeuchi M., Tanakoshi A., Tanaka T., Terpestra P., Tognoni A.,
RA Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vaasarotti A.,
RA Viari A., Wambutt R., Wedler E., Wedler H., Weitzenecker T.,
RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
RA Yoshida K., Yoshikawa H.F., Zumstein E., Yoshikawa H., Danchin A.;
RA "The complete genome sequence of the Gram-positive bacterium Bacillus
RT
```

RT subtilis.";
RL Nature 390:249-256 (1997).
RN [4]
RX SEQUENCE OF 1-56 FROM N.A.
RP MEDLINE=91310584; PubMed=1906866;
RA Grundy F.J., Henkin T.M.;
RT "The rpsd gene, encoding ribosomal protein S4, is autogenously
regulated in *Bacillus subtilis*.";
RL J. Bacteriol. 173:4595-4602 (1991).
RN [5]
RP SEQUENCE OF 1-24.
RX MEDLINE=82219212; PubMed=6806564;
RA Higo K.-I., Otake E., Osawa S.;
RT "Purification and characterization of 30S ribosomal proteins from
Bacillus subtilis: correlation to *Escherichia coli* 30S proteins.";
RL Mol. Gen. Genet. 185:239-244 (1982).
RN [6]
RP VARIANTS.
RX MEDLINE=91035249; PubMed=2121712;
RA Henkin T.M., Chambliss G.H., Grundy F.;
RT "Bacillus subtilis mutants with alterations in ribosomal protein S4.";
RL J. Bacteriol. 172:6380-6385 (1990).
RN [7]
RP CHARACTERIZATION, AND VARIANTS.
RC STRAIN=168;
RA MEDLINE=21382165; PubMed=11489846;
RX Inaoka T., Kasai K., Ochi K.;
RT "Construction of an in vivo nonsense readthrough assay system and
functional analysis of ribosomal proteins S12, S4, and S5 in *Bacillus*
subtilis.";
RL J. Bacteriol. 183:4958-4963 (2001).
CC -!- FUNCTION: One of the primary rRNA binding proteins, it binds
directly to 16S rRNA where it nucleates assembly of the body of
the 30S subunit (By similarity).
CC -!- FUNCTION: With S5 and S12 plays an important role in translational
accuracy; many suppressors of streptomycin-dependent mutants of
protein S12 are found in this protein, some but not all of which
decrease translational accuracy (ram, ribosomal ambiguity
mutations).
CC -!- FUNCTION: S4 represses its own expression; it is not know if this
is at the level of translation or of mRNA stability.
CC -!- SUBUNIT: Part of the 30S ribosomal subunit. Contacts protein S5.
The interaction surface between S4 and S5 is involved in control
of translational fidelity.
CC -!- SIMILARITY: Contains 1 S4 RNA-binding domain.
CC -!- SIMILARITY: Belongs to the S4P family of ribosomal proteins.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC
CC EMBL; M59358; AAA22717.1; -.
CC EMBL; AF008220; AAC00397.1; -.
CC EMBL; S45404; AAB19387.1; -.
CC EMBL; Z99119; CAB14944.1; -.
CC EMBL; M60889; AAA22716.1; -.
CC PIR; A37146; A37146.
CC HSSP; P81288; 1C05.
CC Subtilist; BG10372; rpsd.
CC HAMAP; MF 01306; -; 1.
CC InterPro; IPR001912; Ribosomal_S4.
CC InterPro; IPR005709; Ribosomal_S4_b/o.
CC InterPro; IPR002942; S4.
CC Pfam; PF00163; Ribosomal_S4; 1.
CC Pfam; PF01479; S4; 1.
CC SMART; SM00363; S4; 1.
CC TIGRfams; TIGR01017; rpsd_bact; 1.
CC PROSITE; PS00632; RIBOSOMAL_S4; 1.
CC PROSITE; PS00869; S4; 1.

KW Repressor; Ribosomal protein; RNA-binding; rRNA-binding;
FT Complete proteome.
FT INIT MET 0
FT DOMAIN 91 154 S4 RNA-BINDING.
FT VARIANT 45 45 E -> K (IN RPSD3; SUPPRESSES S12
MUTATION K5SD).
FT VARIANT 74 77 MUTATION K5SD; SUPPRESSES S12
MUTATION K5SD. A RAM MUTATION).
FT VARIANT 77 77 L -> LAGKL (IN RPSD1; SUPPRESSES S12
MUTATION K5SD. A RAM MUTATION).
FT SEQUENCE 199 AA; 22704 MW; 23D669C542E22134 CRC64;
Query Match 6.5%; Score 82.5; DB 1; Length 199;
Best Local Similarity 30.2%; Pred. No. 6.7;
Matches 29; Conservative 14; Mismatches 42; Indels 11; Gaps 3;
QY 7 RADELVFLQGLAESRQAKRLIMAGKVTLTNNSTTIPRLKLEKPKHYPLESICSLGVER 66
DB 92 RLDNVVYKGLARTRQARQLVNHGHILVDGSRVDIPSLVLRFG-----QTIGV-R 141
QY 67 FVSRGAYKLLTALDFEFKIDVKSCICLDAGASTGGFT 102
DB 142 EKSRNLSIIKESVEVNNF-VPEYLTFDAEKLGCTFT 176
RESULT 31
TRME THETN
ID TRME THETN STANDARD; PRT; 460 AA.
AC QSR6K8;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE tRNA modification GTPase trmE.
GN TRME OR THDF OR TTE2796.
OS Thermoanaerobacter tengcongensis.
OC Bacteria; Firmicutes; Clostridia; Thermoanaerobacteriales;
OC Thermoanaerobacteriaceae; Thermoanaerobacter.
OX NCBI_TaxID=119072;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MB4 / JCM 11007;
RX MEDLINE=21992816; PubMed=11997336;
RA Bao Q., Tian Y., Li W., Xu Z., Xuan Z., Hu S., Dong W., Yang J.,
Chen Y., Xue Y., Xu Y., Lai X., Huang L., Dong X., Ma Y., Ling L.,
Tan H., Chen R., Wang J., Yu J., Yang H.;
RT "A complete sequence of *T. tengcongensis* genome.";
RL Genome Res. 12:689-700 (2002).
CC -!- FUNCTION: Exhibits a very high intrinsic GTPase hydrolysis rate.
Involved in the biosynthesis of the hypermodified nucleoside 5-
methylaminomethyl-2-thiouridine, which is found in the wobble
position of some tRNAs (By similarity).
CC -!- SIMILARITY: Belongs to the era/trmE family of GTP-binding
proteins. TrmE subfamily.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC
CC EMBL; AE013217; AAM25900.1; -.
CC HAMAP; MF 00379; -; 1.
CC InterPro; IPR005289; GTP-binding_dom.
CC InterPro; IPR006073; GTPL_ORG.
CC InterPro; IPR002917; MMR_HSR1.
CC InterPro; IPR005225; Small_GTP.
CC InterPro; IPR004520; ThdF.
CC Pfam; PF01926; MMR_HSR1; 1.
CC PRINTS; PR00326; GTPIOB.
CC TIGRfams; TIGR00650; MG442; 1.
CC TIGRfams; TIGR00231; small_GTP; 1.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC -----
DR EMBL; L02415; AAA29898.1; -;
DR EMBL; X05384; CAA28976.1; -;
DR PIR; A48469; A48469.
DR HSP; P08109; 1CKR.
DR InterPro; IPR001023; Hsp70.
DR Pfam; PF00012; HSP70; 1.
DR PRINTS; PR00301; HEATSHOCK70.
DR ProDom; PD000089; Hsp70; 1.
DR PROSITE; PS00297; HSP70_1; 1.
DR PROSITE; PS00329; HSP70_2; 1.
DR PROSITE; PS01036; HSP70_3; 1.
DR ATP-binding; Heat shock.
KW ATP-binding; Heat shock.
FT CONFLICT 69 K -> T (IN REF. 2).
FT CONFLICT 114 A -> P (IN REF. 2).
FT CONFLICT 251 IRS -> NRG (IN REF. 2).
FT CONFLICT 289 T -> I (IN REF. 2).
FT CONFLICT 381 E -> D (IN REF. 2).
FT CONFLICT 564 R -> H (IN REF. 2).
FT CONFLICT 637 AA; A7786D2EB0A11E19 CRC64;
SQ SEQUENCE 637 AA; 69875 MW; 69875 MW; A7786D2EB0A11E19 CRC64;

Query Match 6.5%; Score 82.5; DB 1; Length 637;
Best Local Similarity 21.7%; Pred. No. 25;
Matches 56; Conservative 48; Mismatches 109; Indels 45; Gaps 11;

QY 1 MAKHKVRADELFLQGLAESREQA KRLIMA---GKVLTTNNSTIPIRLKPKGHKPYLES 57
DB 235 MVDFVKEFKYKNDIRSKRALRLRTACERAKRTLSQAQT---NLG-----IDS 284
QY 58 ICSLIGVERFVSRGAYKLLTALDPFK--IDVKSCICLDAGASTGGFTDCLLQHGASKVYA 115
DB 285 LCDGTQFTYVITTRAPPELNA-DLFRGTLDPVEKALDARKDKSQIHDIIVLGGSTRIK 343
QY 116 I-----DVGKG-QLHEKLYTNEQVNIETGVNLTASKDLIPEEVDILTIDVSFISLTIL 169
DB 344 VQKLQDFFNFNGKELINKSINPDEAVYGAAYQAATLSGKCEAVQDILLDLVAPLSGL-- 401
QY 170 PSICRMWKASGIIIALIK-----POEYLPDKIKKGVKVETSLQYEAIVEKIHFHC 219
DB 402 -----ETAGGVMTALIKRNTIPTKTQTFTTYS DN-QPGVL-----IQVFEGERALTKD 450
QY 220 QSELGLIFIGVPSVIKG 237
DB 451 NNLIGKPELSGIPAPRG 468

RESULT 34
KPR2 YEAST
ID KPR2 YEAST STANDARD; PRT; 318 AA.
AC P38620;
DT 01-OCT-1994 (Rel. 30, Created)
DT 01-OCT-1994 (Rel. 30, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Ribose-phosphate pyrophosphokinase 2 (EC 2.7.6.1) (Phosphoribosyl
DE pyrophosphate synthetase 2).
GN PRP52 OR PRS2 OR PRS OR YER099C.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]

RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 44827 / SK02N;
RX MEDLINE=95084630; PubMed=7992503;
RA Carter A.T., Narbad A., Pearson B.M., Beck K.-F., Logghe M.,
RA Contreras R., Schweizer M.;

RT "Phosphoribosylpyrophosphate synthetase (PRS): a new gene family in
RT Saccharomyces cerevisiae.";
RL Yeast 10:1031-1044 (1994).
RN [2]
RP ERATUM.
RA Carter A.T., Narbad A., Pearson B.M., Beck K.-F., Logghe M.,
RA Contreras R., Schweizer M.;
RL Yeast 11:191-191 (1995).
RN [3]
RP SEQUENCE FROM N.A.
RA Gerhardt H., Switzer R.L., Smith J.M., Hove-Jensen B.;
RL Submitted (SEP-1993) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C / AB972;
RX MEDLINE=97313264; PubMed=9169868;
RA Dietrich F.S., Mulligan J.T., Hennessy K.M., Yelton M.A., Allen E.,
RA Araujo R., Aviles E., Berno A., Brennan T., Carpenter J., Chen E.,
RA Cherry J.M., Chung E., Duncan M., Guzman E., Hartzell G.,
RA Hunnicke-Smith S., Hyman R.W., Kayser A., Komp C., Lashkari D., Lew H.,
RA Lin D., Mosedale D., Nakahara K., Namath A., Norgren R., Oefner P.,
RA Oh C., Petel F.X., Roberts D., Sehl P., Schramm S., Shogren T.,
RA Smith V., Taylor P., Wei Y., Botstein D., Davis R.W.;
RL "The nucleotide sequence of Saccharomyces cerevisiae chromosome V.";
RT Nature 367:78-81 (1997).
CC -1- CATALYTIC ACTIVITY: ATP + D-ribose 5-phosphate = AMP + 5-phospho-
CC alpha-D-ribose 1-diphosphate.
CC -1- PATHWAY: Utilized by both the de novo and the salvage pathways by
CC which endogenously formed or exogenously added pyrimidine, purine,
CC or pyridine bases are converted to the corresponding
CC ribonucleoside monophosphates.
CC -1- SIMILARITY: Belongs to the ribose-phosphate pyrophosphokinase
CC family.
CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC -----
DR EMBL; X75075; CAA52969.1; -;
DR EMBL; X74414; CAA52436.1; -;
DR EMBL; U18839; AAB64654.1; -;
DR PIR; S37225; S37225.
DR HSP; P14193; 1DKR.
DR GenOnline; 139179; -;
DR SGD; S0000901; PRS2.
DR InterPro; IPR000842; PRPP_synthetase.
DR InterPro; IPR000836; PRTransferase.
DR InterPro; IPR005946; RibP Ppkin.
DR Pfam; PF00156; Pribosyltran; 1.
DR TIGRfams; TIGR01251; ribp Ppkin; 1.
DR PROSITE; PS00114; PRPP_SYNTHETASE; 1.
DR Nucleotide biosynthesis; Transferase; Kinase; Magnesium;
KW Multigene family.
FT METAL 132 132 MAGNESIUM (POTENTIAL).
FT METAL 134 134 MAGNESIUM (POTENTIAL).
FT METAL 143 143 MAGNESIUM (POTENTIAL).
FT METAL 147 147 MAGNESIUM (POTENTIAL).
SQ SEQUENCE 318 AA; 34765 MW; 89970E98084F5D71 CRC64;

Query Match 6.5%; Score 82; DB 1; Length 318;
Best Local Similarity 29.6%; Pred. No. 12;
Matches 34; Conservative 18; Mismatches 45; Indels 18; Gaps 5;

QY 60 SLIGVERFVSRGAYKLLTALDPFKIDVKSCICLDAGASTGG----FTDCLLQHGASKVYA 115
DB 193 ALIHKRQKANEVSRMLLVGD---VAGKSCLLIDMDADTCGLVACDITMDHGAKEVIA 249
QY 116 I---DVGKGQLHEKLYTNEQVNIETGVNLTASKDLIPEEVDILTIDVSFISLTIL 167

Db 250 IVTHGIFSGAREKL-INSRLSRIVCTN-----TVFVDLDDLDVQDVDSPTI 296

RESULT 35
DNAX_ENTFA STANDARD; PRT; 609 AA.
AC Q835R7;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Chapterone protein dnaK (Heat shock protein 70) (Heat shock 70 kDa protein) (HSP70).
GN DNAK OR EF1308.
OS Enterococcus faecalis (Streptococcus faecalis).
OC Bacteria; Firmicutes; Lactobacillales; Enterococcaceae; Enterococcus.
OX NCBI_TaxID=1351;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=V583 / ATCC 700802;
RX MEDLINE=22550857; PubMed=12663927;
RA Paulsen I.T., Banerjee L., Myers G.S.A., Nelson K.E., Seshadri R., Read T.D., Fouts D.E., Eisen J.A., Gill S.R., Heidelberg J.F., Tettelin H., Dodson R.J., Unayam L., Brinkac L., Beanan M., Daugherty S., DeBoy R.T., Durkin S., Kolonay J., Madupu R., Nelson W., Vamathevan J., Tran B., Upton J., Hansen T., Shetty J., Khouri H., Usterback T., Radune D., Ketchum K.A., Dougherty B.A., Fraser C.M.; "Role of mobile DNA in the evolution of vancomycin-resistant Enterococcus faecalis."
RT Science 299:2071-2074(2003).
RL
CC -!- FUNCTION: Acts as a chaperone (By similarity).
CC -!- INDUCTION: By stress conditions e.g. heat shock (By similarity).
CC -!- SIMILARITY: Belongs to the heat shock protein 70 family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announcement/ or send an email to license@isb-sib.ch).
CC
CC EMBL; AE014951; AAO81100.1; --
DR TIGR; EF1308; --
DR HAMAP; MF_00332; --; 1.
DR InterPro; IPR001023; Hsp70.
DR Pfam; PF00012; HSP70; 1.
DR PRINTS; PR00301; HEATSHOCK70.
DR ProDom; PD000089; HSP70; 1.
DR PROSITE; PS00297; HSP70_1; 1.
DR PROSITE; PS00329; HSP70_2; 1.
DR PROSITE; PS01036; HSP70_3; 1.
KW Chaperone; ATP-binding; Heat shock; Phosphorylation;
KW Complete proteome.
FT MOD_RES 174 174 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
SQ SEQUENCE 609 AA; 65583 MW; 0A28957EA671D4E3 CRC64;

Query Match
Best Local Similarity 21.7%; Pred. No. 26;
Matches 55; Conservative 44; Mismatches 94; Indels 60; Gaps 11;

QY 8 ABELVFLOGLAESRQAQLMAGKVTLTNNSSTIPRLRKPGHYPLESICSLIGVERF 67
DB 224 ANDKMVALQRDKDAEKAKKDLSG-VT-----STQISLPFITAGEAGPLH-----LENN 270
QY 68 VSRGAYKLTA-LDFFKIDVKSCICLDAGASTGGFTDCLLOHGASKAYVAIDVGKGOLHE 125
DB 271 LTRAFKDELTSVLVERTKVPVRQAL-KDAGLNPSIDEIVLGSGSTRIPAV----- 320
QY 126 KLYTNEQVINRGVNLR-T---ASKDLIPEEV-----DILTIDVSFISLTLLPSCI 173
DB 321 -----VBAVKETNKEPKNSENVDEVAMGAIQGVITGDVDKVVLLDVTPLSL 370
QY 174 RWLKASGHIIALIKPFQELYDPDKIKGVVVKETSLEYAEVEKIHFHCOS-----LG 224

Db 371 GIETMGGEVTKLIDRNTIPTSKSQ--VFSTAADNQPAVD--IHVQGERPMADNKTIG 426

QY 225 LIFIGWVPSVIKS 237
DB 427 RFQLTDIPAAPRG 439

RESULT 36
RS4_LACLA STANDARD; PRT; 203 AA.
ID_RS4_LACLA AC Q9C1S2;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE 30S ribosomal protein S4.
GN RPSD OR LL0284.
OS Lactococcus lactis (subsp. lactis) (Streptococcus lactis).
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae; Lactococcus.
OX NCBI_TaxID=1360;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ILL1403;
RX MEDLINE=21235186; PubMed=11337471;
RA Polotini A., Wincker P., Mauger S., Jaillon O., Malarme K., Weissenbach J., Ehrlich S.D., Sorokin A.; "The complete genome sequence of the lactic acid bacterium Lactococcus lactis ssp. lactis ILL1403."
RT Genome Res. 11:731-753(2001).
RL
CC -!- FUNCTION: One of the primary rRNA binding proteins, it binds directly to 16S rRNA where it nucleates assembly of the body of the 30S subunit (By similarity).
CC -!- FUNCTION: With S5 and S12 plays an important role in translational accuracy (By similarity).
CC -!- SUBUNIT: Part of the 30S ribosomal subunit. Contacts protein S5. The interaction surface between S4 and S5 is involved in control of translational fidelity (By similarity).
CC -!- SIMILARITY: Contains 1 S4 RNA-binding domain.
CC -!- SIMILARITY: Belongs to the S4P family of ribosomal proteins.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announcement/ or send an email to license@isb-sib.ch).
CC
CC EMBL; AE006266; AAK04382.1; --
DR PIR; D86660; D86660.
DR HSP; P81288; 1C05.
DR HAMAP; MF_01306; --; 1.
DR InterPro; IPR001912; Ribosomal_S4.
DR InterPro; IPR005709; Ribosomal_S4_b/o.
DR InterPro; IPR002942; S4.
DR Pfam; PF00163; Ribosomal_S4; 1.
DR Pfam; PF01479; S4; 1.
DR SMART; SM00363; S4; 1.
DR TIGRFAMS; TIGR01017; rpsd_bact; 1.
DR PROSITE; PS00632; RIBOSOMAL_S4; 1.
DR PROSITE; PS00889; S4; 1.
KW Ribosomal protein; RNA-binding; Complete proteome.
FT DOMAIN 93 156 S4 RNA-BINDING.
SQ SEQUENCE 203 AA; 23164 MW; 7A7B47CF20887B96 CRC64;

Query Match
Best Local Similarity 32.5%; Pred. No. 8.3;
Matches 26; Conservative 11; Mismatches 32; Indels 11; Gaps 2;

QY 7 RADELVFLOGLAESRQAQLMAGKVTLTNNSSTIPRLRKPGH-----RYP--L 55
DB 94 RLDDNWFERGLATTFRQARQFNHGHLLVDGKVDIPSPRVQGVVISREKSMKVPAIL 153

```
QY 56 ESICSLIGVERFVSRCAYKL 75
Db 154 EAVEATKGRANFVSPDADKL 173

RESULT 37
PRMA THETN
ID PRMA THETN STANDARD; PRT; 309 AA.
AC Q8RB6;
DT 15-MAR-2004 (Rel. 43, Created)
DT 15-MAR-2004 (Rel. 43, Last sequence update)
DE Ribosomal protein L11 methyltransferase (RC 2.1.1.-) (L11 Mtase).
GN PRMA OR TTE0957.
OS Thermoanaerobacter tengcongensis.
OC Bacteria; Firmicutes; Clostridia; Thermoanaerobacteriales;
OC Thermoanaerobacteriaceae; Thermoanaerobacter.
OX NCBI_TaxID=119072;
[1]
RN SEQUENCE FROM N.A.
RC STRAIN=MB4 / JCM 11007;
RX MEDLINE=2192816; PubMed=11997336;
RA Bao Q., Tian Y., Li W., Xu Z., Xuan Z., Hu S., Dong W., Yang J.,
RA Chen Y., Xue Y., Xu Y., Lai X., Huang L., Dong X., Ma Y., Ling L.,
RA Tan H., Chen R., Wang J., Yu J., Yang H.;
RT "A complete sequence of T. tengcongensis genome.";
RL Genome Res. 12:689-700(2002).
CC -1- FUNCTION: Methylates ribosomal protein L11 (By similarity).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -1- SIMILARITY: Belongs to the methyltransferase superfamily. PrmA
family.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL; AE013061; AM24213.1; -.
DR HAMAP; MF 00735; -.
DR InterPro; IPR004498; Ribosomal_Prma.
DR InterPro; IPR000051; SAM bind.
DR TIGRFAMs; TIGR00406; prmA; 1.
KW Transferase; Methyltransferase; Complete proteome.
SQ SEQUENCE 309 AA; 34776 MW; E867A299D1C8508D CRC64;

Query Match 6.4%; Score 81.5; DB 1; Length 309;
Best Local Similarity 25.0%; Pred. No. 13;
Matches 30; Conservative 27; Mismatches 52; Indels 11; Gaps 6;

QY 79 LDPEKIDVK-SCICLDAGATGGTDCCLQHGASKYVAIDVGKQLHEKLYTNEQVINIE 137
Db 164 LQFLEDIVKPCAIVDFVCGSGILSIAASKLGASVYVGADV--DEMAVKI--ARENVKLN 219
QY 138 GV-NLRTASKDLIPE---EVDLIDVFSFSLTLILPSCIRWLKASGIIIA--LIKPOFE 191
Db 220 GLENVEIFQSDLLNFRGKADVIVANIADIRLIIDVLPHLKEEGLFLASGIKDRFE 279

RESULT 38
DNAA TREPA
ID DNAA TREPA STANDARD; PRT; 635 AA.
AC O83246;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Chaperone protein dnaK (Heat shock protein 70) (Heat shock 70 kDa
DE protein) (HSP70).
GN DNAA OR TP0216.
OS Treponema pallidum.
OC Bacteria; Spirochaetes; Spirochaetales; Spirochaetaceae; Treponema.

OX NCBI_TaxID=160;
[1]
RN SEQUENCE FROM N.A.
RC STRAIN=Nichols;
RX MEDLINE=98332770; PubMed=9665876;
RA Fraser C.M., Norris S.J., Weinstock G.M., White O., Sutton G.G.,
RA Dodson R., Gwinn M., Hickey E.K., Clayton R., Ketchum K.A., Peterson J.,
RA Sodergren E., Hardham J.M., McLeod M.P., Salzberg S., Ungerback T.,
RA Khalak H., Richardson D., Howell J.K., Chidambaram M., Ungerback T.,
RA McDonald L., Artach P., Bowman C., Cotton M.D., Fujii C., Garland S.,
RA Hatch B., Horst K., Roberts K., Sandusky M., Weidman J., Smith H.O.,
RA Venter J.C.;
RT "Complete genome sequence of Treponema pallidum, the syphilis
RT spirochete.";
RL Science 281:375-388(1998).
CC -1- FUNCTION: Acts as a chaperone (By similarity).
CC -1- INDUCTION: By stress conditions e.g. heat shock (By similarity).
CC -1- SIMILARITY: Belongs to the heat shock protein 70 family.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL; AE001203; AAC65204.1; -.
DR PIR; F71352; F71352.
DR HSP; P04475; IDG4.
DR TIGR; TP0216; -.
DR HAMAP; MF 00332; -.
DR InterPro; IPR001023; Hsp70.
DR PRINTS; PR00012; HSP70; 1.
DR PROSITE; PR00301; HEATSHOCK70.
DR PRODOM; PD000089; Hsp70; 1.
DR PROSITE; PS00297; HSP70_1; 1.
DR PROSITE; PS00329; HSP70_2; 1.
DR PROSITE; PS01036; HSP70_3; FALSE NEG.
KW Chaperone; ATP-binding; Heat shock; Phosphorylation;
KW Complete proteome.
FT MOD_RES 197 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
SQ SEQUENCE 635 AA; 68041 MW; 11244A863A4504CD CRC64;

Query Match 6.4%; Score 81.5; DB 1; Length 635;
Best Local Similarity 20.8%; Pred. No. 30;
Matches 54; Conservative 45; Mismatches 78; Indels 83; Gaps 14;

QY 14 LQGLAESRQAKRLIMAGKVTLTNNST---IP-LRLEKFGHYPLESICSLIGVERFVS 69
Db 253 LQRLREAAEKA-----KIALSSASSTEINLPFITADANGPKH-----LQRTLS 295
QY 70 RGAVKLLTALDFFKIDVKSC--ICLDAGASTGGTDCCLQHGASKYVAIDVGKQLHEKL 127
Db 296 RSSEFKMTD-DLFEKTEKPCRKALKDAGISADRIDELLVGGSTRMFKV-----A 344
QY 128 YTNREQVINIEGVNLTASKDLIPEV-----DILIDVFSFSLTLILP 170
Db 345 HVIKDVFGEK-----SKGVNPDEAVAIGAAIQGGILGDKVDKVDLLDVTPLSLGI--- 395
QY 171 SCIRWLKASGIIIALIKPQFELYPDKIKGVVETSLQYEAVEKLIHFCQSEGLI---- 226
Db 396 -----ETMGVFTPLISRNTTI---PTRKQVSTAADGQTAVS-IHVLOGERMANQNR 446
QY 227 -----FIGVVPVVIKPGPKG 240
Db 447 TLGNFDLVGIPTA-----PRG 462

RESULT 39
DNAA PSESM
ID DNAA PSESM STANDARD; PRT; 638 AA.
AC Q87WF0;
```

```

10-OCT-2003 (Rel. 42, Created)
10-OCT-2003 (Rel. 42, Last sequence update)
10-OCT-2003 (Rel. 42, Last annotation update)
Chaperone protein dnaK (Heat shock protein 70) (Heat shock 70 kDa
protein) (HSP70).
GN DNAK OR PSPT04505.
OS Pseudomonas syringae (pv. tomato).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OX Pseudomonadaceae; Pseudomonas.
NCBI_TaxID=323;
[1]
RN SEQUENCE FROM N.A.
RC STRAIN=DC3000;
RX MEDLINE=22834015; PubMed=12928499;
RA Buell C.R., Joardar V., Lindeberg M., Selengut J., Paulsen I.T.,
RA Guinn M.L., Dodson R.J., Deboy R.T., Durkin A.S., Kolonay J.F.,
RA Madupu R., Daugherty S., Brinkac L., Beanan M.J., Haft D.H.,
RA Nelson W.C., Davidson T., Zafar N., Zhou L., Liu J., Yuan Q.,
RA Khouri H., Fedorova N., Tran B., Russell D., Berry K., Utterback T.,
RA Van Aken S.E., Feldblyum T.V., D'Ascenzo M., Deng W.-L., Ramos A.R.,
RA Alfano J.R., Catlinhour S., Chatterjee A.K., Delaney T.P.,
RA Lazarowitz S.G., Martin G.B., Schneider D.J., Tang X., Bender C.L.,
RA White O., Fraser C.M., Collier A.;
RT "The complete genome sequence of the Arabidopsis and tomato pathogen
RT Pseudomonas syringae pv. tomato DC3000."
RL Proc. Natl. Acad. Sci. U.S.A. 100:10181-10186(2003).
CC -!- FUNCTION: Acts as a chaperone (By similarity).
CC -!- INDUCTION: By stress conditions e.g. heat shock (By similarity).
CC -!- SIMILARITY: Belongs to the heat shock protein 70 family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AE016872; AA057953.1; -.
DR TIGR; PSPT04505; -.
DR HAMAP; MF_00332; -.
DR InterPro; IPR001023; Hsp70.
DR Pfam; PF00012; HSP70; 1.
DR PRINTS; PR00301; HEATSHOCK70.
DR PROSITE; PS00297; HSP70_1; 1.
DR PROSITE; PS00329; HSP70_2; 1.
DR PROSITE; PS01036; HSP70_3; 1.
KW Chaperone; ATP-binding; Heat shock; Phosphorylation;
KW Complete proteome.
FT MOD_RES 199 199 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
SQ SEQUENCE 638 AA; 68755 MW; E6504919AA678BCC CRC64;

Query Match 6.4%; Score 81.5; DB 1; Length 638;
Best Local Similarity 24.7%; Pred. No. 30;
Matches 65; Conservative 32; Mismatches 89; Indels 77; Gaps 15;

QY 6 VRADELVFLQGLARSREQAKELIMAGKVTLN-----NSTTIPURLEKPGHKYPLESICS 60
Db 252 LKGDFLA-MQLKZAAEKAKIELSSSTQTEVNLPIYTADATGPKHLVVKRSKLSL-- 308
QY 61 LIGVERFVSRYGAYKLLTALDPFKIDVKSICLDAGSTGGFTDCLQHGASKVYADVGK 120
Db 309 ---VEDLVQR-----TIEPCRIALK-----DAGIDIGKINDVILVGGQTRM----- 346
QY 121 GQLHEKLYTNEQVINIGVNWLRASKOLIPREV-----DILPIDVSFI 163
Db 347 -PLVQKLVLT-----EFFEG---KEARKDVNPDEAVAMGAALQGVLAGDVKDVLLDVS-- 395
QY 164 SLTLILPSCIRLWKLKASGIIILIKPQELYPDKTKGVKVKETSQYEAIVEKIHFCCQSE- 222
Db 396 PLTUGIET-----MGVMTALIEKNNTTI-PTK-KSQVFSTADNQNAV--TIHVLOGER 445
QY 223 -----LGLIFIGVVPISVIK 237
```

```

Db 446 KQAGQNSLKGKFDLAEIPAPRG 468
RESULT 40
RS4_BACST
ID RS4_BACST STANDARD; PRT; 199 AA.
AC P81288;
DT 15-DEC-1998 (Rel. 37, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE 30S ribosomal protein S4 (BS5).
GN RPSD.
OS Bacillus stearothermophilus.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Geobacillus.
OX NCBI_TaxID=1422;
RN [1]
RP SEQUENCE.
RX MEDLINE=92110431; PubMed=1764513;
RA Arndt E., Scholzen T., Kromer W., Hatakeyama T., Kimura M.;
RT "Primary structures of ribosomal proteins from the archaeobacterium
RT Halobacterium marismortui and the eubacterium Bacillus
RT stearothermophilus."
RT Biochimie 73:657-668(1991).
RN [2]
RN SEQUENCE FROM N.A., AND X-RAY CRYSTALLOGRAPHY (1.7 ANGSTROMS) OF
RP 41-199.
RX MEDLINE=98372721; PubMed=9707415;
RA Davies C., Gerstner R.B., Draper D.E., Ramakrishnan V., White S.W.;
RT "The crystal structure of ribosomal protein S4 reveals a two-domain
RT molecule with an extensive RNA-binding surface: one domain shows
RT structural homology to the ETS DNA-binding motif."
RL EMBO J. 17:4545-4558(1998).
RN [3]
RN SEQUENCE OF 1-15.
RX STRAIN=10;
RX MEDLINE=75019590; PubMed=4607606;
RA Yaguchi M., Matheson A.T., Visentin L.P.;
RT "Procarvotic ribosomal proteins: N-terminal sequence homologies and
RT structural correspondence of 30 S ribosomal proteins from Escherichia
RT coli and Bacillus stearothermophilus."
RL FEBS Lett. 46:296-300(1974).
RN [4]
RN ISOLATION OF STREPTOMYCIN INDEPENDENT STRAINS.
RC STRAIN=799;
RX MEDLINE=91072295; PubMed=2254291;
RA Schnier J., Gewitz H.S., Behrens S.E., Lee A., Ginther C.,
RA Leighton T.;
RT "Isolation and characterization of Bacillus stearothermophilus 30S and
RT 50S ribosomal protein mutations."
RL J. Bacteriol. 172:7306-7309(1990).
RN [5]
RN BINDING TO RNA.
RX MEDLINE=99425318; PubMed=11401563;
RA Gerstner R.B., Pak Y., Draper D.E.;
RT "Recognition of 16S rRNA by ribosomal protein S4 from Bacillus
RT stearothermophilus."
RL Biochemistry 40:7165-7173(2001).
RN [6]
RN STRUCTURE BY NMR OF 41-198.
RX MEDLINE=20519057; PubMed=11063598;
RA Markus M.A., Gerstner R.B., Draper D.E., Torchia D.A.;
RT "Refining the overall structure and subdomain orientation of
RT ribosomal protein S4 delta41 with dipolar couplings measured by NMR
RT in uniaxial liquid crystalline phases."
RL J. Mol. Biol. 292:375-387(1999).
RN [7]
RN STRUCTURE BY NMR.
RX MEDLINE=20519057; PubMed=11063598;
RA Sayers E.W., Gerstner R.B., Draper D.E., Torchia D.A.;
RT "Structural preordering in the N-terminal region of ribosomal protein
RT S4 revealed by heteronuclear NMR spectroscopy."
RL Biochemistry 39:13602-13613(2000).
```


GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 7, 2004, 16:51:22 ; Search time 58.3721 Seconds
(without alignments)
1356.728 Million cell updates/sec

Title: US-10-009-919A-1

Perfect score: 1271

Sequence: 1 MAKHKVRADELVLFLQGLAES.....PSVIKPGKNGEYLILKKR 251

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- 1: sp_archaea:*
- 2: sp_bacteria:*
- 3: sp_fungi:*
- 4: sp_human:*
- 5: sp_invertebrate:*
- 6: sp_mhc:*
- 7: sp_mammal:*
- 8: sp_organelle:*
- 9: sp_phase:*
- 10: sp_plant:*
- 11: sp_rodent:*
- 12: sp_virus:*
- 13: sp_vertebrate:*
- 14: sp_unclassified:*
- 15: sp_rvirus:*
- 16: sp_bacteriap:*
- 17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	525.5	41.3	271	Q8XJE2	Q8XJE2 clostridium
2	517.5	40.7	258	O66971	O66971 aquifex aeo
3	501.5	39.5	285	Q8RAC4	Q8rac4 thermococcus
4	500.5	39.4	267	Q97HD6	Q97hd6 clostridium
5	500.5	39.4	272	Q894H1	Q894h1 clostridium
6	495.5	39.0	267	Q9X1R2	Q9x1r2 thermotoga
7	448	35.2	267	Q7TUS0	Q7tus0 prochloroc
8	440	34.6	279	Q81M55	Q81m55 bacillus an
9	439.5	34.6	271	Q8DPN1	Q8dpn1 streptococ
10	435.5	34.3	271	Q7U8X7	Q7u8x7 synecococc
11	433	34.1	279	Q818S0	Q818s0 bacillus ce
12	432	34.0	274	P74319	P74319 synecocyst
13	431	33.9	274	Q8V7C0	Q8v7c0 listeria mo
14	429	33.8	262	Q8DIX4	Q8dix4 synecococc
15	426.5	33.6	251	Q92HE9	Q92he9 rickettsia
16	425	33.4	274	Q92BY9	Q92by9 listeria in

17	419.5	33.0	271	16	Q836W3	Q836w3 enterococcu
18	419.5	33.0	272	16	Q88WM4	Q88wm4 lactobacill
19	419.5	33.0	280	16	Q88G54	Q88g54 bifidobacte
20	418.5	32.9	251	16	Q9ZCZ5	Q9zcz5 rickettsia
21	416	32.7	269	16	Q8EQ47	Q8eq47 oceanobacil
22	416	32.7	270	16	Q7TU41	Q7tu41 prochloroco
23	416	32.7	275	16	Q99YX6	Q99yx6 streptococ
24	416	32.7	275	16	Q8P077	Q8p077 streptococ
25	415.5	32.7	276	16	Q8F969	Q8f969 leptospira
26	415	32.7	275	16	Q8K6T0	Q8k6t0 streptococ
27	413	32.5	275	16	Q8DVB2	Q8dvb2 streptococ
28	411	32.3	270	16	Q7VA50	Q7vas0 prochloroco
29	408.5	32.1	245	16	Q89RW3	Q89rw3 bradyrhizob
30	408.5	32.1	272	16	Q9K972	Q9k972 bacillus ha
31	406	31.9	275	16	Q8E6L8	Q8e6l8 streptococ
32	406	31.9	275	16	Q8E163	Q8e163 streptococ
33	403.5	31.7	281	2	Q9L3Q7	Q9l3q7 mycobacteri
34	400	31.5	266	16	Q8REA5	Q8rea5 fusobacteri
35	391	30.8	260	2	Q9F6Y3	Q9f6y3 chloroflexu
36	387.5	30.5	245	16	Q92RJ2	Q92rj2 rhizobium m
37	384.5	30.3	269	16	O05676	O05676 mycobacteri
38	376.5	29.6	268	16	Q50760	Q50760 mycobacteri
39	376.5	29.6	268	16	Q7VEU4	Q7veu4 mycobacteri
40	374.5	29.5	301	10	Q9LSV5	Q9lsv5 arabidopsis
41	373	29.3	246	16	Q8EWO3	Q8ewg3 mycoplasma
42	372.5	29.3	250	16	O8UHE6	O8uhe6 agrobactera
43	365	28.7	242	16	Q8YIE4	Q8yye4 anabaena sp
44	364	28.6	268	2	Q49898	Q49898 mycobacteri
45	363	28.6	243	16	Q9CH80	Q9ch80 lactococcus

ALIGNMENTS

RESULT 1

Q8XJE2 ID Q8XJE2 PRELIMINARY; PRT; 271 AA.
AC Q8XJE2;
DT 01-MAR-2002 (TREMBLrel. 20, Created)
DT 01-MAR-2002 (TREMBLrel. 20, Last sequence update)
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE Probable hemolysin.
GN HLYD OR CPE1818.
OS Clostridium perfringens.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OC NCBI_TaxID=1502;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=13 / Type A;
RX MEDLINE=21664373; PubMed=11792842;
RA Shimizu T., Ohtani K., Hirakawa H., Ohshima K., Yamashita A.,
Shiba T., Ogasawara N., Hattori M., Kuhara S., Hayashi H.;
RT "Complete genome sequence of Clostridium perfringens, an anaerobic
flesh-eater.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:996-1001(2002).
DR ENBL; AP003191; BAB81524.1; -;
DR GO; GO:0003723; rRNA binding; IEA.
DR GO; GO:0015070; F:toxin activity; IEA.
DR InterPro; IPR002877; RrmJ_FtsJ.
DR InterPro; IPR002942; S4.
DR InterPro; IPR004538; Tly.
DR Pfam; PF01728; FtsJ; 1.
DR Pfam; PF01479; S4; 1.
DR SMART; SM00363; S4; 1.
DR TIGRFAMs; TIGR00478; tly; 1.
DR PROSITE; PS50889; S4; 1.
KW Complete proteome.
SQ SEQUENCE 271 AA; 30269 MW; 0142CE2A326C692C CRC64;

Query Match

Best Local Similarity 49.6%; Score 525.5; DB 16; Length 271;

Matches 125; Conservative 36; Mismatches 74; Indels 17; Gaps 5;

QY 5 KVRADLVFLOGLAHSREQAELIMAGKVTLTNNSTTIPRLKPGHKYPLESICSLIGV 64
 Db 8 KERDLVLLVEQGLAHSREKAKYIMAGVFIQEK-----RVDKAGEKVPVSSNIIFRGE 61
 QY 65 E-RFVSRGAYKLLTALDFFKIDVKSCICLDAGASTGGFTDCLLQHGASKVYVADVGKGL 123
 Db 62 KLPFVSRGGFKLDKAVKSGFDLKGKRCCLDAGASTGGFTDCLQNDASKVFSIDVGYGQF 121
 QY 124 HEKLYTNEQVINIEGVNLTASKOLIPREVDILTIDVSFISLTLLPSCIRWLKASGIIII 183
 Db 122 AWKLRVDPVWCNVRTNRYVTPQIGELCDFASIDVSFISLTLLPVLINLNDNGEVM 181
 QY 184 ALIKPOFELYPKIK-KGVVVKETSLOEAVEKIIHFCQSE---LGLIFIGVVPVSIKGP 238
 Db 182 ALIKPOFEAGREKVGKGVVREASTHKEVKKIVDFALSHKLNILGLDF-----SPIKGP 236
 QY 239 KGNQEVLYLYKK 250
 Db 237 EGNIEVLYLYKK 248

RESULT 2
 ID C66971 PRELIMINARY; PRT; 258 AA.
 AC C66971;
 DT 01-AUG-1998 (TrEMBLrel. 07, Created)
 DT 01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Hemolysin.
 GN TLY OR AQ_773.
 OS Aquifex aeolicus.
 OC Bacteria; Aquificae; Aquificales; Aquificaceae; Aquifex.
 OC NCBI_TaxID=63363;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=VF5;
 RX MEDLINE=98136666; PubMed=9537320;
 RA Deckert G., Warren P.V., Gaasterland T., Young W.G., Lenox A.L.,
 RA Graham D.E., Overbeek R., Snead M.A., Keller M., Aujay M., Huber R.,
 RA Feldman R.A., Short J.M., Olson G.J., Swanson R.V.,
 RA "The complete genome of the hyperthermophilic bacterium Aquifex
 RT aeolicus.";
 RL Nature 392:353-358(1998).
 DR EMBL; AE00707; AAC06935.1; .
 DR PIR; H70367.
 DR GO; GO:0003723; F:RNA binding; IEA.
 DR GO; GO:0008757; F:8-adenosylmethionine-dependent methyltransf. . .; IEA.
 DR GO; GO:0015070; F:toxin activity; IEA.
 DR InterPro; IPR002877; RrmJ_FtsJ.
 DR InterPro; IPR002942; S4.
 DR InterPro; IPR000051; SAM_bind.
 DR InterPro; IPR004538; Tly.
 DR Pfam; PF01728; FtsJ; 1.
 DR Pfam; PF01479; S4; 1.
 DR SMART; SM00363; S4; 1.
 DR TIGRFAMs; TIGR00478; tly; 1.
 DR SMART; SM00363; S4; 1.
 DR TIGRFAMs; TIGR00478; tly; 1.
 DR PROSITE; PS50889; S4; 1.
 KW Complete proteome.
 SQ SEQUENCE 258 AA; 176C1E87A309B6F7 CRC64;

Query Match 40.7%; Score 517.5; DB 16; Length 258;
 Best Local Similarity 43.6%; Pred. No. 4.5e-39;
 Matches 105; Conservative 57; Mismatches 72; Indels 7; Gaps 3;
 QY 6 VRADLVFLOGLAHSREQAELIMAGKVTLTNNSTTIPRLKPGHKYPLESICSLIGVE 65
 Db 1 MRLDKYLTGKGVFSREKAQAVIMAGV-LVNGKV-----VDKPGYRLKNGKEVVEKLP 54
 QY 66 RFVSRGAYKLLTALDFFKIDVKSCICLDAGASTGGFTDCLLQHGASKVYVADVGKGLHE 125
 Db 55 KYVSRGGKLEWAKRFSLDLKDVKVLDVGSSTGGFTDCLQHGAKVYVADVGRGQWDY 114

QY 126 KLYTNEQVINIEGVNLTASKDLIPREVDILTIDVSFISLTLLPSCIRWLKASGIIIIAL 185
 Db 115 KLRQPRVVLVYESTDARELSEHVPKVDLITCDVSFISSTKVLNPNVFKLXEDGLLLVL 174
 QY 186 IKPOFELYPKIK-KGVVVKETSLOEAVEKIIHFCQSELGLIFIGVVPVSIKGPKNQYXL 245
 Db 175 VAPQELCPKVKGVREKEHKEALQKVNFLK-ENGFRILGVIKSPKPGTKGNEEFF 233
 QY 246 I 246
 Db 234 V 234

RESULT 3
 ID Q8RAC4 PRELIMINARY; PRT; 265 AA.
 AC Q8RAC4;
 DT 01-JUN-2002 (TrEMBLrel. 21, Created)
 DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Predicted rRNA methylase.
 GN TTE1299.
 OS Thermoanaerobacter tengcongensis.
 OC Bacteria; Firmicutes; Clostridia; Thermoanaerobacteriales;
 OC Thermoanaerobacteriaceae; Thermoanaerobacter.
 OC NCBI_TaxID=119072;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=MB4 / JCM 11007;
 RX MEDLINE=21992816; PubMed=11997336;
 RA Bao Q., Tian Y., Li W., Xu Z., Xuan Z., Hu S., Dong W., Yang J.,
 RA Chen Y., Xue Y., Xu Y., Lai X., Huang L., Dong X., Ma Y., Ling L.,
 RA Tan H., Chen R., Wang J., Yu J., Yang H.;
 RA "A complete sequence of T. tengcongensis genome.";
 RT Genome Res. 12:689-700(2002).
 RL EMBL; AE013091; AAM24523.1; .
 DR GO; GO:0008168; F:methyltransferase activity; IEA.
 DR GO; GO:0003723; F:RNA binding; IEA.
 DR GO; GO:0015070; F:toxin activity; IEA.
 DR InterPro; IPR002877; RrmJ_FtsJ.
 DR InterPro; IPR002942; S4.
 DR InterPro; IPR004538; Tly.
 DR Pfam; PF01728; FtsJ; 1.
 DR Pfam; PF01479; S4; 1.
 DR SMART; SM00363; S4; 1.
 DR TIGRFAMs; TIGR00478; tly; 1.
 DR PROSITE; PS50889; S4; 1.
 DR Methyltransferase; Complete proteome.
 KW SEQUENCE 265 AA; 29549 MW; 266D29F3E71FBA9D CRC64;

Query Match 39.5%; Score 501.5; DB 16; Length 265;
 Best Local Similarity 47.6%; Pred. No. 1.4e-37;
 Matches 118; Conservative 39; Mismatches 80; Indels 11; Gaps 5;
 QY 5 KVRADLVFLOGLAHSREQAELIMAGKVTLTNNSTTIPRLKPGHKYPLESICSLIGV 64
 Db 2 KERDLVLLVRRGFFSSREKAAIMAGEVVDGK-----RAEKAGEMWEDSKIEVKN 55
 QY 65 E-RFVSRGAYKLLTALDFFKIDVKSCICLDAGASTGGFTDCLLQHGASKVYVADVGKGL 123
 Db 56 SLFVSVSRGGLKLEKALQFLGIDVRGKIALDVGASTGGFTDCLLKHGAQKVVAVDVGQGL 115
 QY 124 HEKLYTNEQVINIEGVNLTASKDLIPREVDILTIDVSFISLTLLPSCIRWLKASGIIII 183
 Db 116 HWSLRNDPRVVMKTNIRFL--NALPEMVDIITIDVSFISLTIVVPAADKFLKSEGEIV 173
 QY 184 ALIKPOFELYPKIK-KGVVVKETSLOEAVEKIIHFCQSELGLIFIGVVPVSIKGPKNQ 242
 Db 174 ALIKPOFEAGREKVGKGVVREKVDVHKEVLEKIKLEKN-INYGVCCITYSPIKAGBN 232
 QY 243 EYLYLYKK 250
 Db 233 EYLYLYKK 240

```
RESULT 4
Q97HD6
ID Q97HD6 PRELIMINARY; PRT; 267 AA.
AC Q97HD6;
DT 01-OCT-2001 (TrEMBLrel. 18, Created)
DT 01-OCT-2001 (TrEMBLrel. 18, Last sequence update)
DE 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Predicted rRNA methylase, YQXC B. subtilis ortholog.
GN CAC2076
OS Clostridium acetobutylicum.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1488;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 824 / DSM 792 / VOM B-1787;
RX MEDLINE=21359325; PubMed=11466286;
RA Noelling J., Breton G., Omelchenko M.V., Makarova K.S., Zeng Q.,
RA Gibson R., Lee H.M., Dubois J., Qiu D., Hitti J., Wolf Y.I.,
RA Tatusov R.L., Sabathe F., Doucette-Stamm L., Soucaille P., Daly M.J.,
RA Bennett G.N., Koonin E.V., Smith D.R.;
RA "Genome sequence and comparative analysis of the solvent-producing
RT bacterium Clostridium acetobutylicum.";
RL J. Bacteriol. 183:4823-4838(2001).
DR EMBL; AE007710; AAK80035.1; -.
DR PIR; H97155; H97155.
DR GO; GO:0008168; F:methyltransferase activity; IEA.
DR GO; GO:0003723; F:RNA binding; IEA.
DR GO; GO:0015070; F:toxin activity; IEA.
DR InterPro; IPR002877; RsmJ_FtsJ.
DR InterPro; IPR002942; S4.
DR Pfam; PF01728; FtsJ; 1.
DR Pfam; PF01479; S4; 1.
DR SMART; SMO0363; S4; 1.
DR TIGRFAMs; TIGR00478; tly; 1.
DR PROSITE; PS50889; S4; 1.
KW Methyltransferase; Complete proteome.
SQ SEQUENCE 267 AA; 29324 MW; 7B389BC667C609C3 CRC64;

Query Match 39.4%; Score 500.5; DB 16; Length 267;
Best Local Similarity 44.7%; Pred. No. 1.7e-37;
Matches 113; Conservative 49; Mismatches 82; Indels 9; Gaps 4;

QY 1 MAKHKVRADELVFLQGLAESREQAKRLIMAGKVTLTNNSTIPIRLKPGHKYPLESICS 60
DB 1 MSENKERDLVLLVEKGFESREKARASIMAGEIYVDD-----LRIDKCGQKVKVSKVE 54
QY 61 LIGVER-FVSRGAYKLLTALDFFKIDVKSCICLDAGASTGGFTDCLLQHGASKVYVADVG 119
DB 55 FRGEKMPYVSRGGYKLEAKISFGLGKDKVCFDYGASTGGFTDCLMONGASKVFAIDVG 114
QY 120 KGQHEKLYNEQVINIEGVNLRITASKDLIPEEVDIITDVSFISLTLLIPSCIRWLKAS 179
DB 115 YGQFAWKLRITDPRVVCMERINVRVTPEDIGFCNFPASIDVFSLSLKKVIPVWNLKDD 174
QY 180 GIITALLKPOFELYPDKI-KKGUVKETSQYAEVEKLIHFCQSLGLIFGVVPSVIKGP 238
DB 175 GEIVALLKPOFEAGREKVGKRVREPTHEVINTIVDFLK-EMKLSILGITYSPIKGP 233
QY 239 KGNQEYLILYKKR 251
DB 234 EGNIEYLIVFSKK 246

RESULT 5
Q894H1
ID Q894H1 PRELIMINARY; PRT; 272 AA.
AC Q894H1;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
```

```
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hemolysin.
GN CTC01574.
OS Clostridium tetani.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1513;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Massachusetts / E88;
RX MEDLINE=22457253; PubMed=12552129;
RA Brueggemann H., Baeumer S., Fricke W.F., Wiezer A., Liesegang H.,
RA Decker I., Herzberg C., Martinez-Arias R., Merkl R., Henne A.,
RA Gottschalk G.;
RT "The genome sequence of Clostridium tetani, the causative agent of
RT tetanus disease.";
RL Proc. Natl. Acad. Sci. U.S.A. 100:1316-1321(2003).
DR EMBL; AE015941; AAO36121.1; -.
DR GO; GO:0003723; F:RNA binding; IEA.
DR InterPro; IPR002877; RsmJ_FtsJ.
DR InterPro; IPR002942; S4.
DR Pfam; PF01728; FtsJ; 1.
DR Pfam; PF01479; S4; 1.
DR PROSITE; PS50889; S4; 1.
KW Complete proteome.
SQ SEQUENCE 272 AA; 30911 MW; 0664D095C30C1A7E CRC64;

Query Match 39.4%; Score 500.5; DB 16; Length 272;
Best Local Similarity 46.4%; Pred. No. 1.7e-37;
Matches 117; Conservative 42; Mismatches 84; Indels 9; Gaps 4;

QY 1 MAKHKVRADELVFLQGLAESREQAKRLIMAGKVTLTNNSTIPIRLKPGHKYPLESICS 60
DB 1 MAEKKLRDLVLLVEKEIFTGRERARSSIMAGEIFVDG-----IRKDKCGKINVDAXIE 54
QY 61 LIG-VERFVSRGAYKLLTALDFFKIDVKSCICLDAGASTGGFTDCLLQHGASKVYVADVG 119
DB 55 FKGDTPYVSRGGYKLEKAKDFNVDLKDKVCLDIGASTGGFTDCLMONGASKVYVADVG 114
QY 120 KGQHEKLYNEQVINIEGVNLRITASKDLIPEEVDIITDVSFISLTLLIPSCIRWLKAS 179
DB 115 YGQLAWKLRITNPVKVCMERTNFRHKEHIGEVADFPASIDVFSLSLEKLIIPAINILNDK 174
QY 180 GIITALLKPOFELYPDKI-KKGUVKETSQYAEVEKLIHFCQSLGLIFGVVPSVIKGP 238
DB 175 GEIVALLKPOFEAGREKVGKRVREKSTHIEVIEKLVDFAKT-LDLKIKGINYSPIKGP 233
QY 239 KGNQEYLILYKK 250
DB 234 EGNIEYLIVYTK 245

RESULT 6
Q9X1R2
ID Q9X1R2 PRELIMINARY; PRT; 267 AA.
AC Q9X1R2;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hemolysin.
GN TW1576.
OS Thermotoga maritima.
OC Bacteria; Thermotogae; Thermotogales; Thermotogaceae; Thermotoga.
OX NCBI_TaxID=2336;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MSB8 / DSM 3109;
RX MEDLINE=99287316; PubMed=10360571;
RA Nelson K.E., Clayton R.A., Gill S.R., Gwinn M.L., Dodson R.J.,
RA Haft D.H., Hickey E.K., Peterson J.D., Nelson W.C., Ketchum K.A.,
RA McDonald L., Utterback T.R., Malek J.A., Linher K.D., Garrett M.M.,
RA Stewart A.M., Cotton M.D., Pratt M.S., Phillips C.A., Richardson D.,
RA Heidelberg J., Sutton G.G., Fleischmann R.D., Eisen J.A., White O.,
```

```
RA Salzberg S.L., Smith H.O., Venter J.C., Fraser C.M.;
RT "Evidence for lateral gene transfer between Archaea and Bacteria from
RT Genome sequence of Thermotoga maritima.";
RL Nature 399:323-329(1999).
DR EMBL; AE001802; AAD36643.1; -.
DR PIR; A72238; A72238.
DR TIGR; TW1576; -.
DR GO; GO:0003723; F:RNA binding; IEA.
DR GO; GO:0015070; F:toxin activity; IEA.
DR InterPro; IPR002877; RrmJ_FtsJ.
DR InterPro; IPR002942; S4.
DR InterPro; IPR004538; Tly.
DR Pfam; PF01728; FtsJ; 1.
DR Pfam; PF01479; S4; 1.
DR SMART; SM00363; S4; 1.
DR TIGRFAMS; TIGR00478; tly; 1.
DR PROSITE; PS0889; S4; 1.
DR Complete proteome.
SQ SEQUENCE 267 AA; 30045 MW; F47AAEE2F2AE6092 CRC64;

Query Match 39.0%; Score 495.5; DB 16; Length 267;
Best Local Similarity 47.2%; Pred. No. 4.8e-37;
Matches 116; Conservative 37; Mismatches 86; Indels 7; Gaps 3;

QY 5 KVRADLVFLQGLAESREQAKRLIMAGKVTLTNNSTTPIRLLEKPGHKYPLESICSLIGV 64
DB 4 KKELDQLVLRGLVESREKAKVILAGKV-LVNGE-----RVTKASKLVDPEDANVELLEE 57
QY 65 ERVSRGAYKLLTALDFFKIDVKSICLDAGASTGGFTDCLLOHGASKYVAIDVGKQLH 124
DB 58 PKYVSRGGYKLESAPESFKIDVSGKVCADIGASTGGFTDCLLOHGARKKVVAVDVGQQLH 117
QY 125 EKLTYNEQVINIGVNLRTASKDLIPREVLDITDVSFISLTILPSCIRWLKASGIIIA 184
DB 118 WKLRNDPRVVMVKWARYNLPDQGEKVDVTCIDVSFISLKKIIPAIKILKNIGDAVL 177
QY 185 LIKQPELYPDKIKKGVKETSLOQEAVEKIIHFCQSELGLIFIGVVPVSVIKGPKGNQEV 244
DB 178 LVKQPEAPRKFKVKGIVKDPGVHLEVLB-IRKSLIENGFFVVKGCFCFSIKIGTEGNI 236
QY 245 LIVLKK 250
DB 237 FFWVKK 242

RESULT 7
Q7TUSO PRELIMINARY; PRT; 267 AA.
AC Q7TUSO;
DT 01-OCT-2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE FtsJ cell division protein:S4 domain:hemolysin A.
GN PWT1482.
OS Prochlorococcus marinus (strain MIT 9313).
OC Bacteria; Cyanobacteria; Prochlorophytes; Prochlorococcaceae;
OC Prochlorococcus.
OC NCBI_TaxID=74547;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=22825698; PubMed=12917642;
RA Rocap G., Larimer F.W., Lamerdin J., Malfatti S., Chain P.,
RA Algren N.A., Arellano A., Coleman M., Hauser L., Hess W.R.,
RA Johnson Z.I., Land M., Lindell D., Post A.F., Regala W., Shah M.,
RA Shaw S.L., Steglich C., Sullivan M.B., Ting C.S., Tolonen A.,
RA Webb E.A., Zinser E.R., Chistom S.W.;
RT "Genome divergence in two Prochlorococcus ecotypes reflects oceanic
RT niche differentiation.";
RL Nature 424:1042-1047(2003).
DR EMBL; BX572099; CAE21657.1; -.
DR Cell division; Complete proteome.
SQ SEQUENCE 267 AA; 29226 MW; 6A26ID2762A1335C CRC64;
```

```
Query Match 35.2%; Score 448; DB 16; Length 267;
Best Local Similarity 43.8%; Pred. No. 1.1e-32;
Matches 110; Conservative 42; Mismatches 87; Indels 12; Gaps 6;

QY 4 HVRADLVFLQGLAESREQAKRLIMAGKVTLTNNSTTPIRLLEKPGHKYPLESICSLIG 63
DB 3 HKQRDLQLLMKGLASSRHOAQLTRAGKVRNGQL-----LDKPGHEVSQELEQVEQ 57
QY 64 VERFVSRGAYKLLTALDFFKIDVKSICLDAGASTGGFTDCLLOHGASKYVAIDVGKQL 123
DB 58 PRFVSRGGEKLLGALREFPLNVEGRVCLDGGISTGGFTDCLLOHGAAARVYDVGQGT 117
QY 124 HEKLYTNEOVINIGVNLRTASKDLI---PEEVDILTI-DVSFISLTILPSCIRWLKAS 179
DB 118 AWWLRNDRSVLRTNLRTLSPLDQLYGAEELPLSLAVADLAFISLRLVPAIKLKKPD 177
QY 180 -GIITALIKPOPELYPDKI-KGVVKETSLOQEAVEKIIHFCQSELGLIFIGVVPVSVIKG 237
DB 178 HSEAVLLVRPQPEVGRERVKGKGVVRDALAHVDALXSVIDTSRS-LGWPKGLIASPITG 236
QY 238 PKGNQEVLYL 248
DB 237 PAGNHEYLLWL 247

RESULT 8
Q8IM55 PRELIMINARY; PRT; 279 AA.
AC Q8IM55;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hemolysin A.
GN BA4399.
OS Bacillus anthracis (strain Ames).
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OC NCBI_TaxID=198094;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=22608414; PubMed=12721629;
RA Read T.D., Peterson S.N., Tourasse N., Baillie L.W., Paulsen I.T.,
RA Nelson K.E., Tettelin H., Fouts D.E., Eisen J.A., Gill S.R.,
RA Holtzapple E.K., Okstad O.A., Helgason E., Rilstone J., Wu M.,
RA Kolonay J.F., Beanan M.J., Dodson R.J., Brinkac L.M., Gwinn M.,
RA DeBoy R.T., Madpu R., Daugherty S.C., Durkin A.S., Haft D.H.,
RA Nelson W.C., Peterson J.D., Pop M., Khouri H.M., Radune D.,
RA Benton J.L., Mahmoud Y., Jiang L., Hance I.R., Weidman J.F.,
RA Berry K.J., Plaut R.D., Wolf A.M., Watkins K.L., Nierman W.C.,
RA Hazen A., Cline R., Redmond C., Thwaite J.E., White O., Salzberg S.L.,
RA Thomson B., Friedlander A.M., Koehler T.M., Hanna P.C., Kolsto A.-B.,
RA Fraser C.M.;
RT "The genome sequence of Bacillus anthracis Ames and comparison to
RT closely related bacteria.";
RL Nature 423:81-86(2003).
DR EMBL; AE017037; AAP28114.1; -.
DR TIGR; BA4399; -.
DR GO; GO:0003723; F:RNA binding; IEA.
DR GO; GO:0015070; F:toxin activity; IEA.
DR InterPro; IPR002877; RrmJ_FtsJ.
DR InterPro; IPR002942; S4.
DR InterPro; IPR004538; Tly.
DR Pfam; PF01728; FtsJ; 1.
DR Pfam; PF01479; S4; 1.
DR SMART; SM00363; S4; 1.
DR TIGRFAMS; TIGR00478; tly; 1.
DR PROSITE; PS0889; S4; 1.
DR Complete proteome.
SQ SEQUENCE 279 AA; 31138 MW; A448DC82A567585F CRC64;

Query Match 34.6%; Score 440; DB 16; Length 279;
Best Local Similarity 44.0%; Pred. No. 6.1e-32;
Matches 111; Conservative 39; Mismatches 92; Indels 10; Gaps 5;
```


QY 1 MAKHKVRADELVFLQGLAESREQAKRLIMAGKVTLTNNSTTIPLRLKPGHKYPLESICS 60
 Db 1 MSAKKERVDVLLVERGLIETREKAKRAIMAGLVVANE-----NRDKPGKIPQDTEIT 54
 QY 61 LIG-VERFVSRGAYKLTALDFFKIDVKSCICLDAGASTGGTDCLLQHGASKYVAIDVG 119
 Db 55 VKGQVMPYVRGGYKLEKALFTHLDIDQKVMIDIGSGTGGTDCALQNGAKLSYALDVG 114
 QY 120 KGQHEKLYTNEQVINTEGVNLR-TASKDLIPEEVDILTIDVSFISLTILPSCIRMLKA 178
 Db 115 YNQLAWKLRODERVVMERINFRVTPADLERGHPQFASIDVSFISLKLILPVLKTMMP 174
 QY 179 SGIILAIKPOFELYPKI-KKGVVKETSLOYEAVEKIIHFQCSLGLIFIGVVPVSIK 237
 Db 175 NGDVAALIKPOFEAGREQVGKGIVRDRKVEAVEMIVDFAIKE-GYDVEGLTFSPITG 233
 QY 238 PKGNOEYLIYK 249
 Db 234 GDGNIEFLIHLK 245

RESULT 9

Q8DPN1 Q8DPN1 PRELIMINARY; PRT; 271 AA.
 AC Q8DPN1; 23, Created)
 DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)
 DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
 DE Conserved hypothetical protein.
 GN SPRI086.
 OS Streptococcus pneumoniae (strain ATCC BAA-255 / R6).
 OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
 OC Streptococcus.
 CX NCBI_TaxID=171101;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=21429245; PubMed=11544234;
 RA Hoskins J., Albom W.E. Jr., Arnold J., Blaszcak L.C., Burgett S.,
 RA DeHoff B.S., Estrem S.T., Fritz L., Fu D.-J., Fuller W., Geringer C.,
 RA Gilmour R., Glass J.S., Khoja H., Kraft A.R., Lagace R.E.,
 RA LeBlanc D.J., Lee L.N., Lefkowitz E.J., Lu J., Matsushima P.,
 RA Mahren S.M., McMeney M., Mcleaster K., Mundy C.W., Nicas T.I.,
 RA Norris F.H., O'Gara M., Peery R.B., Robertson G.T., Rockey P.,
 RA Sun P.-M., Winkler M.E., Yang Y., Young-Bellido M., Zhao G.,
 RA Zook C.A., Baltz R.H., Jaskunas S.R., Rostock P.R. Jr., Skatrud P.L.,
 RA Glass J.I.;
 RT "Genome of the bacterium Streptococcus pneumoniae strain R6.";
 RL J. Bacteriol. 183:5709-5717(2001).
 DR EMBL; AEO08481; AAK99889.1; -.
 DR PIR; E98007; E98007.
 DR GO; GO:0003723; F:RNA binding; IEA.
 DR GO; GO:0015070; F:toxin activity; IEA.
 DR InterPro; IPR02877; RmJ_FtsG.
 DR InterPro; IPR02942; S4.
 DR InterPro; IPR004538; Tly.
 DR Pfam; PF01728; FtsG; 1.
 DR Pfam; PF01479; S4; 1.
 DR SMART; SM00363; S4; 1.
 DR TIGRFAMs; TIGR00478; tly; 1.
 DR PROSITE; PS00889; S4; 1.
 DR Hypothetical protein; Complete proteome.
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 271 AA; 29869 MW; 58C27C6C43AA99A5 CRC64;

Query Match 34.6%; Score 439.5; DB 16; Length 271;
 Best Local Similarity 42.5%; Pred. No. 6.5e-32;
 Matches 107; Conservative 41; Mismatches 95; Indels 9; Gaps 5;

QY 1 MAKHKVRADELVFLQGLAESREQAKRLIMAGKVTLTNNSTTIPLRLKPGHKYPLESICS 60
 Db 1 MAKERV-DVLAYKQGLFETREQAKRGVAGLVAVLNGE----RFDKPGKIPDDTELK 54
 QY 61 LIGVE-RFVSRGAYKLTALDFFKIDVKSCICLDAGASTGGTDCLLQHGASKYVAIDVG 119

Db 55 LKGEKLYVSRGGKLEKALQVFDLSVDGATTIDIGASTGGTDMQLQNSAKLVFAVDVG 114
 QY 120 KGQHEKLYTNEQVINTEGVNLR-TASKDLIPEEVDILTIDVSFISLTILPSCIRMLKAS 179
 Db 115 TNQLAWKLRODPRVSMEQNFRVAETDPEQEFSPASIDVSFISLILPALHRVLADQ 174
 QY 180 GIILAIKPOFELYPKI-KKGVVKETSLOYEAVEKIIHFQCSLGLIFIGVVPVSIK 238
 Db 175 GQVVALVKPOFEAGREQIGKNGIIRDAKHQNVLESYTAMA-VEAGSFVLDDSPICGG 233
 QY 239 KGNOEYLIYK 250
 Db 234 HGNIEFLAYLKK 245

RESULT 10

Q7UBX7 Q7UBX7 PRELIMINARY; PRT; 271 AA.
 AC Q7UBX7;
 DT 01-OCT-2003 (TREMBlrel. 25, Created)
 DT 01-OCT-2003 (TREMBlrel. 25, Last sequence update)
 DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
 DE Hemolysin-like protein.
 GN SYNW0482.
 OS Synechococcus sp. (strain WH8102).
 OC Bacteria; Cyanobacteria; Chroococcales; Synechococcus.
 CX NCBI_TaxID=84588;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=22825697; PubMed=12917641;
 RA Palenik B., Brahmsha B., Larimer F.W., Land M., Hauser L., Chain P.,
 RA Lamerdin J., Regalia W., Allen E.E., McCarren J., Paulsen I.,
 RA Dufresne A., Partensky F., Webb E.A., Waterbury J.;
 RT "The genome of a motile marine Synechococcus.";
 RL Nature 424:1037-1042(2003).
 RL EMBL; BX569690; CAE06997.1; -.
 KW Complete proteome.
 SQ SEQUENCE 271 AA; 29030 MW; CD3BC1E6B46EBB4A CRC64;

Query Match 34.3%; Score 435.5; DB 16; Length 271;
 Best Local Similarity 41.6%; Pred. No. 1.5e-31;
 Matches 111; Conservative 40; Mismatches 73; Indels 43; Gaps 8;

QY 5 KVRADLVFLQGLAESREQAKRLIMAGKVTLTNNSTTIPLRLKPGHKYP 54
 Db 4 KQRLDLELLSGLVWSRQQAQLIRAGKVRDAGTLLDKPGTEVAALALELVEQP 58
 QY 55 LESICSLIGVERFVSRGAYKLTALDFFKIDVKSCICLDAGASTGGTDCLLQHGASKY 114
 Db 59 -----PRFVSRGGKLAGLKAFFIATEGRVCLDGGISTGGTDCLLQHGASRV 108
 QY 115 AIDVGKQLHEKLYTNEQVINTEGVNLR-TASKDLIPEEVDILTIDVSFISLT 166
 Db 109 GVDVGGTQTAWSLRTDVRVLRRTNLR----HLQPEQLYGAEDPWPFSLATVDSFISLR 164
 QY 167 LIPLSCIRMLKASGII----IALIKPOFELYPKI-KKGVVKETSLOYEAVEKIIHFCS 221
 Db 165 LVLPALRRLLOPAGGLCPALVLVKQFQVGKDRGVGGVVRDPVAHRDAIE-LVMAAA 223
 QY 222 ELGLIFIGVVPVSIKPGKNOEYLIYL 248
 Db 224 ELGWYPOGIVASPTPGAGNHEVVLWL 250

RESULT 11

Q818S0 Q818S0 PRELIMINARY; PRT; 279 AA.
 AC Q818S0;
 DT 01-JUN-2003 (TREMBlrel. 24, Created)
 DT 01-JUN-2003 (TREMBlrel. 24, Last sequence update)
 DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
 DE Putative cell adhesion protein (contact hemolysin TlyA-related).
 GN BC4175.


```
DR EMBL; AL591978; CAC99444.1; -.
DR PIR; AF1245; AF1245.
DR ListList; LMO01366; -.
DR GO; GO:0003723; F:RNA binding; IEA.
DR GO; GO:0015070; F:toxin activity; IEA.
DR InterPro; IPR002877; RnmJ_FtsJ.
DR InterPro; IPR002942; S4.
DR Pfam; PF01728; FtsJ; 1.
DR SMART; SM00363; S4; 1.
DR TIGRFAMs; TIGR00478; tly; 1.
DR PROSITE; PS00889; S4; 1.
DR KW Hypothetical protein; Complete proteome.
DR SEQUENCE 274 AA; 30520 MW; 094A3BC96F567DF1 CRC64;

Query Match      33.9%; Score 431; DB 16; Length 274;
Best Local Similarity 43.3%; Pred. No. 3.9e-31;
Matches 109; Conservative 36; Mismatches 97; Indels 10; Gaps 5;

Qy 1 MAKHKVRADELVFLQGLAESREQAQRLIMAGKVTLTNNSTTIPLRLKPGHKYPLESICS 60
Db 1 MTIKKERADILLVQGLFETREKAKRAIMAGIVRKER-----RVDPGGEKIPIDSELQ 54

Qy 61 LIGVER-FVSRGAYKLLTALDFKIDVKSCICLDAGASTGGFTDCLLQHGASKVVAIDVG 119
Db 55 VKGQMPVVSRRGGLKLEKALQVDFEVKDKLMDIGASTGGFTDCLQNGARHSYALDVG 114

Qy 120 KGQHEKLYTNEQVINIEGVNLR-TASKDLIPEE---VDILTIDVSFISLTILPSCIRML 178
Db 115 YNQLAWLRNDRVTVMERTRFHVKPADFAEGLADFATIDVSFISLTILPVLRTLVLT 174

Qy 179 SGIIIALIKPQFELYPKI-KKGVVKETSLOYEAVEKIIHFCQSELGLIFGVVPSVIK 237
Db 175 GGDVMTLIKQFAGRGVQVKGKGIIRDPVAVHESVVEHLVQFALDN-GVDLMGLDYSPTG 233

Qy 238 PKGNQOEYLIYK 249
Db 234 GEGNIEFTAHLK 245

RESULT 14
Q8DIX4
ID Q8DIX4 PRELIMINARY; PRT; 262 AA.
AC Q8DIX4;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hemolysin-like protein.
GN TLL1457.
OS Synechococcus elongatus (Thermosynechococcus elongatus).
OC Bacteria; Cyanobacteria; Chroococcales; Synechococcus.
OX NCBI_TaxID=32045;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BP-1;
RX MEDLINE=2225144; PubMed=12240834;
RA Nakamura Y., Kaneko T., Sato S., Ikeuchi M., Katoh H., Sasamoto S.,
RA Watanabe A., Iriguchi M., Kawashima K., Kimura T., Kishida Y.,
RA Kiyokawa C., Kohara M., Matsumoto M., Matsuno A., Nakazaki N.,
RA Shimpo S., Sugimoto M., Takeuchi C., Yamada M., Tabata S.;
RT "Complete genome structure of the thermophilic cyanobacterium
RL Thermosynechococcus elongatus BP-1.";
RL EMBL; AP005374; BAC09009.1; -.
DR GO; GO:0003723; F:RNA binding; IEA.
DR GO; GO:0015070; F:toxin activity; IEA.
DR InterPro; IPR002877; RnmJ_FtsJ.
DR InterPro; IPR002942; S4.
DR InterPro; IPR004538; Tly.
DR Pfam; PF01728; FtsJ; 1.
DR SMART; SM00363; S4; 1.
DR KW Complete proteome.
DR SEQUENCE 274 AA; 30520 MW; 094A3BC96F567DF1 CRC64;

Query Match      33.9%; Score 431; DB 16; Length 274;
Best Local Similarity 43.3%; Pred. No. 3.9e-31;
Matches 109; Conservative 36; Mismatches 97; Indels 10; Gaps 5;

Qy 1 MAKHKVRADELVFLQGLAESREQAQRLIMAGKVTLTNNSTTIPLRLKPGHKYPLESICS 60
Db 1 MTIKKERADILLVQGLFETREKAKRAIMAGIVRKER-----RVDPGGEKIPIDSELQ 54

Qy 61 LIGVER-FVSRGAYKLLTALDFKIDVKSCICLDAGASTGGFTDCLLQHGASKVVAIDVG 119
Db 55 VKGQMPVVSRRGGLKLEKALQVDFEVKDKLMDIGASTGGFTDCLQNGARHSYALDVG 114

Qy 120 KGQHEKLYTNEQVINIEGVNLR-TASKDLIPEE---VDILTIDVSFISLTILPSCIRML 178
Db 115 YNQLAWLRNDRVTVMERTRFHVKPADFAEGLADFATIDVSFISLTILPVLRTLVLT 174

Qy 179 SGIIIALIKPQFELYPKI-KKGVVKETSLOYEAVEKIIHFCQSELGLIFGVVPSVIK 237
Db 175 GGDVMTLIKQFAGRGVQVKGKGIIRDPVAVHESVVEHLVQFALDN-GVDLMGLDYSPTG 233

Qy 238 PKGNQOEYLIYK 249
Db 234 GEGNIEFTAHLK 245

RESULT 14
Q8DIX4
ID Q8DIX4 PRELIMINARY; PRT; 262 AA.
AC Q8DIX4;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hemolysin-like protein.
GN TLL1457.
OS Synechococcus elongatus (Thermosynechococcus elongatus).
OC Bacteria; Cyanobacteria; Chroococcales; Synechococcus.
OX NCBI_TaxID=32045;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BP-1;
RX MEDLINE=2225144; PubMed=12240834;
RA Nakamura Y., Kaneko T., Sato S., Ikeuchi M., Katoh H., Sasamoto S.,
RA Watanabe A., Iriguchi M., Kawashima K., Kimura T., Kishida Y.,
RA Kiyokawa C., Kohara M., Matsumoto M., Matsuno A., Nakazaki N.,
RA Shimpo S., Sugimoto M., Takeuchi C., Yamada M., Tabata S.;
RT "Complete genome structure of the thermophilic cyanobacterium
RL Thermosynechococcus elongatus BP-1.";
RL EMBL; AP005374; BAC09009.1; -.
DR GO; GO:0003723; F:RNA binding; IEA.
DR GO; GO:0015070; F:toxin activity; IEA.
DR InterPro; IPR002877; RnmJ_FtsJ.
DR InterPro; IPR002942; S4.
DR InterPro; IPR004538; Tly.
DR Pfam; PF01728; FtsJ; 1.
DR SMART; SM00363; S4; 1.
DR KW Complete proteome.
DR SEQUENCE 274 AA; 30520 MW; 094A3BC96F567DF1 CRC64;
```

```
DR TIGRFAMs; TIGR00478; tly; 1.
DR PROSITE; PS00889; S4; 1.
DR KW Complete proteome.
DR SEQUENCE 262 AA; 28607 MW; 1B5291F08133A27B CRC64;

Query Match      33.8%; Score 429; DB 16; Length 262;
Best Local Similarity 43.3%; Pred. No. 5.7e-31;
Matches 109; Conservative 39; Mismatches 92; Indels 12; Gaps 6;

Qy 1 MAKHKVRADELVFLQGLAESREQAQRLIMAGKVTLTNNSTTIPLRLKPGHKYPLESICS 60
Db 1 MTPRKQRLDSLVERHLCSRQQAQWIRAGAVVNH---IP--IDKPGTLVAVDATIQ 54

Qy 61 LIGVERFVSRCAYKLLTALDFKIDVKSCICLDAGASTGGFTDCLLQHGASKVVAIDVGK 120
Db 55 VKAKPAYSVRGGEKLAHALGVFPVVRDRVCLDGGISTGFTDCLLQAGAQLVYIDVGY 114

Qy 121 GQLHEKLYTNEQVINIEGVNLR-TASKDLIPEE---VDILTIDVSFISLTILPSCIRML 176
Db 115 GQVWKLRQDPRLLIRBETNLRVITPADLYTDDGPRDGVVDVSFISLTIKVLPALWELL 174

Qy 177 KASGIIIALIKPQFELYPKI-KKGVVKETSLOYEAVEKIIHFCQSELGLIFGVVPSVI 235
Db 175 LPPRELIALIKPQFEGVRDRGLKGVVDRDAKARQEAVDQVI-TAAALGWOCYGVTPSPI 233

Qy 236 KGPKNQOEYLIY 247
Db 234 LGPAGNQOEFLAH 245

RESULT 15
Q92HE9
ID Q92HE9 PRELIMINARY; PRT; 251 AA.
AC Q92HE9;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hemolysin.
GN TLYA OR RC0822.
OS Rickettsia conorii.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
OC Rickettsiaceae; Rickettsiae; Rickettsia.
OX NCBI_TaxID=781;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Malish 7;
RX MEDLINE=21442074; PubMed=11557893;
RA Ogata H., Audic S., Renesto-Audiffren P., Fournier P.-E., Barbe V.,
RA Samson D., Roux V., Cossart P., Weissenbach J., Claverie J.-M.,
RA Raoult D.;
RT "Mechanisms of evolution in Rickettsia conorii and R. prowazekii.";
RL Science 293:2093-2098(2001).
DR EMBL; AE008637; AAL03360.1; -.
DR PIR; F97802; F97802.
DR GO; GO:0003723; F:RNA binding; IEA.
DR GO; GO:0015070; F:toxin activity; IEA.
DR InterPro; IPR002877; RnmJ_FtsJ.
DR InterPro; IPR002942; S4.
DR InterPro; IPR004538; Tly.
DR Pfam; PF01728; FtsJ; 1.
DR SMART; SM00363; S4; 1.
DR TIGRFAMs; TIGR00478; tly; 1.
DR PROSITE; PS00889; S4; 1.
DR KW Complete proteome.
DR SEQUENCE 251 AA; 28383 MW; 4AEA1CB69C1752BB CRC64;

Query Match      33.6%; Score 426.5; DB 16; Length 251;
Best Local Similarity 37.7%; Pred. No. 9.1e-31;
Matches 98; Conservative 47; Mismatches 86; Indels 29; Gaps 4;

Qy 5 KVRADLVFLQGLAESREQAQRLIMAGKVTLTNNSTTIPLRLKPGHK 52
Db 3 KIRLDEVLLQKGFVTITTIARSLIIQGVNKHKEQLIKFGIQVINDTEIKVL--PQHN 60
```

QY 53 YPELICSILGVERFVSRYAYKLLTALDPFKIDVKSCICLDAGASTGGFTDCLLOHGASK 112
Db 61 Y-----VSRGALKLIALDYFKIDPENLVICIDIGSGTGGTEVLLERKAKL 106
QY 113 VYVIDVGKQOLHEKLYTNEQVNIIEGVNLTASKOLIPBEVDILTIDVSFISLTILPSC 172
Db 107 IFADVGVGELHPKLRDNPHKVLEKTNARYLTDKQIITKPLIVCDASFISLTILPTV 166
QY 173 IRLWKASGIIILIPQFELYDPDKYK-CVVKETSLOYEAVEKLIHFQCSQELGLIFIGV 231
Db 167 LNLVKEDOMLIALIKPQFEVEKHEVQGVIKNPLLHQVKDKIKDWLEKHEHFKIFGII 226
QY 232 PSVIGKPGKNOEVLILYKKR 251
Db 227 ASPLGAKNQEFLLCGKKR 246

RESULT 16

Q92BY9 PRELIMINARY; PRT; 274 AA.
AC Q92BY9; (TREMBlrel. 19, Created)
DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE Hypothetical protein lin1403.
GN LIN1403.
OS *Listeria innocua*.
OC Bacteria; Firmicutes; Bacillales; Listeriaceae; Listeria.
OX NCBI_TaxID=1642;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CLIP 11262 / Serovar 6a;
RX MEDLINE=21597279; PubMed=11679669;
RA Glaser P., Frangeul L., Buchrieser C., Rueniok C., Amend A.,
BAQUERO F., Berche P., Bloeker H., Brandt P., Chakraborty T.,
Charbit A., Chetoui F., Couve E., de Daruvar A., Dehoux P.,
Domann E., Dominguez-Bernal G., Duchaud E., Durand L., Dussurget O.,
Entian K.-D., Fsihi H., Garcia-del Portillo F., Garrido P.,
Gautier L., Goebel W., Gomez-Lopez N., Hain T., Hauf J., Jackson D.,
Jones L.-M., Kaerst U., Kretz J., Kuhn M., Kunst F., Kurapat G.,
Madueno E., Maitournam A., Mata Vicente J., Ng E., Nedjari H.,
Nordstiek G., Novella S., de Pablos B., Perez-Diaz J.-C., Purcell R.,
Rammel B., Rose M., Schluster T., Simoes N., Tierrez A.,
Vazquez-Roland J.-A., Voss H., Wehland J., Cossart P.;
RT "Comparative genomics of *Listeria* species";
RL Science 294:849-852(2001).
DR EMBL; AL596168; CAC96634.1; -.
DR PIR; AB1608; AB1608.
DR ListList; LIN01403; -.
DR GO; GO:0003723; F:RNA binding; IEA.
DR GO; GO:0015070; F:Toxin activity; IEA.
DR InterPro; IPR002877; RrmJ_FtsJ.
DR InterPro; IPR002942; S4.
DR InterPro; IPR004538; Tly.
DR Pfam; PF01728; FtsJ; 1.
DR SMART; SM00363; S4; 1.
DR TIGR; TIGR00478; tly; 1.
DR TIGR; TIGR00478; tly; 1.
DR PROSITE; PS00889; S4; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 274 AA; 30424 MW; 5FA61FD0E77BFAB9 CRC64;

Query Match 33.4%; Score 425; DB 16; Length 274;
Best Local Similarity 43.7%; Pred. No. 1.4e-30;
Matches 110; Conservative 33; Mismatches 99; Indels 10; Gaps 5;
QY 1 MAKHVRADLVFLQGLAESREQAKRLIMAGKVTLTNNSTTIPRLKPKGHKYPLESTICS 60
Db 1 MTIKKERADILLVQGLFETREKAKRALMAGIVYRKE-----RVDKPGKIPADSELQ 54
QY 61 LIGVER-FVSRGAYKLLTALDPFKIDVKSCICLDAGASTGGFTDCLLOHGASKVYVDVG 119

Db 55 VVGKOMPVYVSRGGLKLEKALQVFNEDVKDKMLMDIGASTGGFTDCLQNGARHSYALDVG 114
QY 120 KQQLHEKLYTNEQVNIIEGVNLTASKOLIPBEVDILTIDVSFISLTILPSCIRWLKA 178
Db 115 YNLAWKLRDNERVTVMTNFRHVTVPADFTGLAEAFATIDVSFISLTILPVLRLVLT 174
QY 179 SGIIIALIKPQELYPDKI-KKGVVKETSLOYEAVEKLIHFQCSQELGLIFIGVPSVIGK 237
Db 175 GGDVMTLTKPQFEAGREQVKGKGIIRDPAVHAENVENTIALFALDN-GYDLMGLDFSPITG 233
QY 238 PKGNOEVLILYK 249
Db 234 GEGNIEFIAHLK 245

RESULT 17

Q836W3 PRELIMINARY; PRT; 271 AA.
AC Q836W3;
DT 01-JUN-2003 (TREMBlrel. 24, Created)
DT 01-JUN-2003 (TREMBlrel. 24, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE Hemolysin A.
GN TLYA OR EF0982.
OS *Enterococcus faecalis* (Streptococcus faecalis).
OC Bacteria; Firmicutes; Lactobacillales; Enterococcaceae; Enterococcus.
OX NCBI_TaxID=1351;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=V583 / ATCC 700802;
RX MEDLINE=22550857; PubMed=12663927;
RA Paulsen I.T., Banerjee L., Myers G.S.A., Nelson K.E., Seshadri R.,
Read T.D., Fouts D.E., Eisen J.A., Gill S.R., Heidelberg J.F.,
Tettelin H., Dodson R.J., Umayam L., Brinkac L., Beanan M.,
Daugherty S., DeBoy R.T., Durkin S., Kolonay J., Madupu R., Nelson W.,
Vamathevan J., Tran B., Upton J., Hansen T., Shetty J., Khouri H.,
Uttarback T., Radune D., Ketchum K.A., Dougherty B.A., Fraser C.M.;
RT "Role of mobile DNA in the evolution of vancomycin-resistant
Enterococcus faecalis";
RL Science 299:2071-2074(2003).
DR EMBL; AE016950; AA080788.1; -.
DR TIGR; EF0982; -.
DR GO; GO:0003723; F:RNA binding; IEA.
DR GO; GO:0015070; F:Toxin activity; IEA.
DR InterPro; IPR002877; RrmJ_FtsJ.
DR InterPro; IPR002942; S4.
DR InterPro; IPR004538; Tly.
DR Pfam; PF01728; FtsJ; 1.
DR Pfam; PF01479; S4; 1.
DR TIGR; TIGR00478; tly; 1.
DR TIGR; TIGR00478; tly; 1.
KW Complete proteome.
SQ SEQUENCE 271 AA; 30062 MW; 777A0AA910F6B511 CRC64;

Query Match 33.0%; Score 419.5; DB 16; Length 271;
Best Local Similarity 44.9%; Pred. No. 4.4e-30;
Matches 111; Conservative 33; Mismatches 94; Indels 9; Gaps 5;
QY 5 KVRADLVFLQGLAESREQAKRLIMAGKVTLTNNSTTIPRLKPKGHKYPLESTICSIG- 63
Db 3 KERVDVLAFNQGLFETREKAKRSVMAGLVYNDKNE-----RLDKPGKISVETPLHTKGQ 57
QY 64 VERFVSRGAYKLLTALDPFKIDVKSCICLDAGASTGGFTDCLLOHGASKVYVDVGKQL 123
Db 58 VMPVYVSRGGLKLEKALNVFAINVQGTMLDLSGSGTGGFTDVALQNGARLSYALDVGYNQL 117
QY 124 HEKLYTNEQVNIIEGVNLTASKOLIPBEVDILTIDVSFISLTILPSCIRWLKASGII 182
Db 118 AWKIQDERVVVWERTNFRYKSKPEDFTGVPDIATIDVSFISLTILPPLHLILKKGVS 177
QY 183 IALIKPQFELYDPDKI-KKGVVKETSLOYEAVEKLIHFQCSQELGLIFIGVPSVIGKPKGN 241
Db 178 VALIKPQFEAGKEFVGKGIIVRDPETHQMVVEITRFAMNN-GYDVKNLDFSPITGGEGN 236

GN	BL1048.
OS	Bifidobacterium longum.
OC	Bacteria; Actinobacteria; Actinobacteridae; Bifidobacteriales;
OC	Bifidobacteriaceae; Bifidobacterium.
OX	NCBI_taxID=216816;
RN	[1]
RP	SEQUENCE FROM N.A.
RC	STRAIN=NCC 2705;
RX	MEDLINE=22294977; PubMed=12381787;
RA	Schell M.A., Karmirantzou M., Snel B., Vilanova D., Berger B.,
RA	Pessi G., Zwielen M.-C., Desiere F., Bork P., Delley M.,
RA	Pridmore R.D., Arigoni F.,
RT	"The genome sequence of Bifidobacterium longum reflects its adaptation
RT	to the human gastrointestinal tract.";
RT	Proc. Natl. Acad. Sci. U.S.A. 99:14422-14427(2002).
DR	EMBL; AE014728; AAN24854.1; -.
DR	GO; GO:0003723; F:RNA binding; IEA.
DR	GO; GO:0015070; F:toxin activity; IEA.
DR	InterPro; IPR002877; RnmJ_FtsJ.
DR	InterPro; IPR002942; S4.
DR	InterPro; IPR004538; Tly.
DR	Pfam; PF01728; FtsJ; 1.
DR	Pfam; PF01479; S4; 1.
DR	SMART; SM00363; S4; 1.
DR	TIGRFAMS; TIGR00478; tly; 1.
DR	PROSITE; PSS0889; S4; 1.
KW	Complete proteome.
SQ	SEQUENCE 280 AA; 29817 MW; 7DA44D7D5BA1EE4 CRC64;
Query Match 33.0%; Score 419.5; DB 16; Length 280;	
Best Local Similarity 40.2%; Pred. No. 4.5e-30;	
Matches 100; Conservative 42; Mismatches 90; Indels 17; Gaps 5;	
QY	7 RADELVFLQGLAESREQAKRLIMAGKVTLTNNSTTIPRLKPKGHKYPLESICSLIGVER 66
Db	37 RLDMLVASGLVESRAKARLIKAGHVRVDGETTIKPSFMVKAGH-----SELAVDK 88
QY	67 ---FVSRGAYKLLTALDFFK---IDVKSCICLDAGASTGGFTDCLLQHGASKVYVADVG 119
Db	89 GDYVSRGAYKLLGAFETPAGNLTPGEGLECLDYGASTGGFTDVLRLGGAKVVALDVG 148
QY	120 KQGLHEKLYTNEQVINIEGVNLTASKDLIPEEVDILTIDVSFISLTILPSCIRWLKAS 179
Db	149 HQGLDPRIANDEHVIEMSGVNIREFEADDLPYPAMIVSDVSFISLTIVPVIARIAAPG 208
QY	180 GIILIKPOFELYPDKI-KKGVVKETSLEYEAWEKIIHFQCOSELGLIFTGVVPSVIKGP 238
Db	209 AQIVLLVFPQFVGRAGLKGKNGIVEDPALRERALHDVV-ACAEQHGLDVTADSPITGT 267
QY	239 KGNQEVLLY 247
Db	268 HGNREYLLY 276
RESULT 20	
ID	Q9ZCZ5 PRELIMINARY; PRT; 251 AA.
AC	Q9ZCZ5;
DT	01-MAY-1999 (TREMBlrel. 10, Created)
DT	01-MAY-1999 (TREMBlrel. 10, Last sequence update)
DT	01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE	Hemolysin (TLYA).
GN	RP555.
OS	Rickettsia prowazekii.
OC	Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
OC	Rickettsiaceae; Rickettsiae; Rickettsia.
OX	NCBI_TaxID=782;
RN	[1]
RP	SEQUENCE FROM N.A.
RC	STRAIN=Madrid E;
RX	MEDLINE=99039499; PubMed=9823893;
RA	Andersson S.G.E., Zomorodipour A., Andersson J.O.,
RA	Sicheritz-Ponten T., Alsmark U.C.M., Podowski R.M., Naeslund A.K.,

QY	242 QEYLIVL 248
Db	237 IEFLAHL 243
RESULT 18	
Q88WM4	
ID	Q88WM4 PRELIMINARY; PRT; 272 AA.
AC	Q88WM4;
DT	01-JUN-2003 (TREMBlrel. 24, Created)
DT	01-JUN-2003 (TREMBlrel. 24, Last sequence update)
DT	01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE	Hemolysin homolog.
GN	LP_1603.
OS	Lactobacillus plantarum.
OC	Bacteria; Firmicutes; Lactobacillales; Lactobacillaceae;
OC	Lactobacillus.
OX	NCBI_TaxID=1590;
RN	[1]
RP	SEQUENCE FROM N.A.
RC	STRAIN=NCIMB 8826 / WCFS1;
RX	MEDLINE=22480296; PubMed=12566566;
RA	Kleerebezem M., Boekhorst J., van Kranenburg R., Molenaar D.,
RA	Kuipers O.P., Leer R., Tarchini R., Peters S.A., Sandbrink H.M.,
RA	Fliers M.W.E.J., Stiekema W., Klein Lankhorst R.M., Bron P.A.,
RA	Hoffer S.M., Nierop Groot M.N., Kerkhoven R., De Vries M., Ursing B.,
RA	De Vos W.M., Siezen R.J.;
RA	"Complete genome sequence of Lactobacillus plantarum WCFS1.";
RL	Proc. Natl. Acad. Sci. U.S.A. 100:1990-1995(2003).
DR	EMBL; AL935256; CAD64044.1; -.
DR	GO; GO:0003723; F:RNA binding; IEA.
DR	InterPro; IPR002877; RnmJ_FtsJ.
DR	InterPro; IPR002942; S4.
DR	Pfam; PF01728; FtsJ; 1.
DR	Pfam; PF01479; S4; 1.
DR	PROSITE; PSS0889; S4; 1.
KW	Complete proteome.
SQ	SEQUENCE 272 AA; 29797 MW; 72AC156857D6234C CRC64;
Query Match 33.0%; Score 419.5; DB 16; Length 272;	
Best Local Similarity 41.9%; Pred. No. 4.4e-30;	
Matches 104; Conservative 44; Mismatches 91; Indels 9; Gaps 5;	
QY	5 KVRADELVFLQGLAESREQAKRLIMAGKVTLTNNSTTIPRLKPKGHKYPLESICSLIGV 64
Db	3 KERVDVLLVQQGLDFREKANGRAVMAEILGVNEE-----RLDKPGMKIPVATELHLKKG 57
QY	65 ER-FVSRGAYKLLTALDFFKIDYKSCICLDAGASTGGFTDCLLQHGASKVYVADVGKQL 123
Db	58 PMFYVSRGGGLKLEKALKSFEIDVTDKVLIDIGSGTGFTDVLQNGAKMSYALDVGSNQL 117
QY	124 HEKLYTNEQVINIEGVNLTAS-KDLIPEEVDILTIDVSFISLTILPSCIRWLKASGII 182
Db	118 VWKLQDPVNVVMEHTNFRYSKLADFTQQGFNFASIDVSVFISLHLILPPLHAIENGGSV 177
QY	183 IALIKPOFELYPDKI-KKGVVKETSLEYEAWEKIIHFQCOSELGLIFTGVVPSVIKPGKN 241
Db	178 VALIKQFAGRENVGHGIVRDPVAVHQAVITDIVEFAQAN-GYNVLGLDYSIKGEGN 236
QY	242 QEYLIVLK 249
Db	237 IEFLIHQ 244
RESULT 19	
Q8G5G4	
ID	Q8G5G4 PRELIMINARY; PRT; 280 AA.
AC	Q8G5G4;
DT	01-MAR-2003 (TREMBlrel. 23, Created)
DT	01-MAR-2003 (TREMBlrel. 23, Last sequence update)
DT	01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE	Hemolysin-like protein with S4 domain found in bacteria and
DE	plants.

EA	Eriksson A.-S., Winkler H.H., Kurland C.G.;
RT	"The genome sequence of Rickettsia prowazekii and the origin of
RI	Rickettsia mitochondria."
RL	Nature 396:133-140(1998).
DR	EMBL; AJ235272; CAJ15004.1; --.
DR	PIR; B71660; B71660.
DR	GO; GO:0003723; F:RNA binding; IEA.
DR	InterPro; IPR002877; RimJ_FtsJ.
DR	InterPro; IPR002942; S4.
DR	Pfam; PF01728; FtsJ; 1.
DR	SMART; SM00363; S4; 1.
DR	PROSITE; PS00889; S4; 1.
KW	Complete proteome.
SQ	SEQUENCE 251 AA; 28315 MW; 1684B544D526120A CRC64;
Query Match	32.9%; Score 418.5; DB 16; Length 251;
Best Local Similarity	37.0%; Pred.No.4.9e-30;
Matches	95; Conservative 49; Mismatches 88; Indels 25; Gaps 3;
Qy	5 KVRADLVFLVGLAESREQAKRLIMAGKV-----TITNNTTIPLRLEKPGHKYP 54
Dd	3 KIRLDYLLQKLGLVTDAIARSLLIQGVHNKHEKLIKSGIKVNRHDPDIKKVLPOHNY- 61
Qy	55 LESICSLIGVERFVSRCAYKLLITALDFPKIDVKSCICLDAGASGGTDCLLQHGASKVY 114
Dd	62 -----VSRGALKLITALDYFKIDPQNLCVICDIGSGSTGGTEVLFFERKAELIF 108
Qy	115 AIDVGKQLHEKYLTNEQVINIEGWNLRTASKOLIPREVDILITDVFSISITLILPSCIR 174
Dd	109 ADVGVGELHSKLFNFNQIKVEKTARYITDKQITTKPDLLIVCDASFISLITILPTPLN 168
Qy	175 WLKASGIILAIKPQELYDPDKKK-GVWKETSIOYBAVEKIIHFQOSELGLIFIGVVPS 233
Dd	169 LAKEDCILLIALKPQEVKNVENGGIINPLLHQVCCKIKMWLBQEHFHFIQFIIGIAS 228
Qy	234 VIKGPKNORYLYLVKK 250
Dd	229 PILGTGNKEFLICGR 245
RESULT 21	
Q8EQ47	PRELIMINARY; PRT; 269 AA.
ID Q8EQ47	PRELIMINARY; PRT; 269 AA.
AC Q8EQ47	PRELIMINARY; PRT; 269 AA.
DT 01-MAR-2003	(TrEMBLrel. 23, Created)
DT 01-MAR-2003	(TrEMBLrel. 23, Last sequence update)
DE 01-OCT-2003	(TrEMBLrel. 25, Last annotation update)
DE Hemolysin.	
GN O81876.	
OS Oceanobacillus ihayensis.	
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Oceanobacillus.	
OX NCBI_TaxID=182710;	
ON [1]	
RP SEQUENCE FROM N.A.	
RC STRAIN=HTE631 / DSM 14371 / JCM 11309;	
RX MEDLINE=22220767; PubMed=12235376;	
RA Takami H., Takaki Y., Uchiyama I.;	
RT "Genome sequence of Oceanobacillus ihayensis isolated from the Ihaya	
RT Ridge and its unexpected adaptive capabilities to extreme	
RT environments."	
RL Nucleic Acids Res. 30:3927-3935(2002).	
DR EMBL; AF004599; BAC13832.1; --.	
DR GO; GO:0015070; F:toxin activity; IEA.	
DR InterPro; IPR002877; RimJ_FtsJ.	
DR InterPro; IPR004538; Tly.	
DR Pfam; PF01728; FtsJ; 1.	
DR TIGRFAMS; TIGR00478; tly; 1.	
KW Complete proteome.	
SQ SEQUENCE 269 AA; 30539 MW; 85CC793ECDFC8BE5 CRC64;	
Query Match	32.7%; Score 416; DB 16; Length 269;
Best Local Similarity	39.7%; Pred.No..9e-30;
Matches	104; Conservative 50; Mismatches 72; Indels 36; Gaps 8;

RESULT 23					
C	D	Q99YX6	PRELIMINARY;	PRT;	275 AA.
C	D	Q99YX6;			
I	T	01-JUN-2001	(TrEMBLrel. 17, Created)		
I	T	01-JUN-2001	(TrEMBLrel. 17, Last sequence update)		
T	T	01-OCT-2003	(TrEMBLrel. 25, Last annotation update)		
E	N	Putative hemolysin.			
N	N	HLYA1 OR SPY1497.			
S	S	Streptococcus pyogenes.			
S	S	Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;			
C	C	Streptococcus			
X	X	NCBI_TaxID=1314;			
N	N	[1]			
P	P	SEQUENCE FROM N.A.			
C	C	STRAIN-SF370 / ATCC 700294 / Serotype M1;			
X	X	MEDLINE=21192684; PubMed=11296296;			
A	A	Ferratti J.J., McShan W.M., Ajdic D.J., Savic D.J., Savic G., Lyon K.,			
A	A	Primeaux C., Sezate S., Suvorov A.N., Kenton S., Lai H.S., Lin S.P.,			
A	A	Qian Y., Jia H.G., Najjar F.Z., Ren Q., Zhu H., Song L., White J.,			
A	A	Yuan X., Clifton S.W., Roe B.A., McLaughlin R.;			
L	L	"Complete genome sequence of an MI strain of Streptococcus pyogenes.";			
T	T	Proc. Natl Acad. Sci. U.S.A. 98:4658-4663(2001).			
R	R	ENBL; AS006583; AKK34296.1; --			
R	R	GO; GO:0003723; FRNA binding; IEA.			
R	R	GO; GO:001070; F.toxin activity; IEA.			
R	R	InterPro; IPR002877; RrmJ_FtsJ.			
R	R	InterPro; IPR002942; S4.			
R	R	InterPro; IPR004538; Tly.			
R	R	Pfam; PF01728; FtsJ; 1.			
R	R	Pfam; PF01479; S4; 1.			
R	R	SMART; SM00363; S4; 1.			
R	R	TIGRFAMS; TIGR00478; tly; 1.			
R	R	PROSITE; PS50889; S4; 1.			
W	W	Complete proteome.			
Q	Q	SEQUENCE 275 AA; 30440 MW; CA936893EAC91E08 CRC64;			
Query Match 32.7%; Score 416; DB 16; Length 275;					
Best Local Similarity 43.5%; Pred. No. 9.3e-30;					
Matches 110; Conservative 37; Mismatches 90; Indels 16; Gaps 6;					
a	y	5 KVPADLVFLQGLAESREQAKLIMAGKYTLINNSTTIPRLLEKPKGHKYPLESICSLIGV 64			
b	b	3 KERVDVLAYKQGLFETREQAKRGWMAGLVVSINGO----RYDKPKDKIDDGTEKLKGE 58			
y	y	65 E-RFVSRGAYKLLTALDFFFKIDVKSCICLDAGASTGGFTDCLLQHGAASKYVAIDVGKQL 123			
b	b	59 KLKTVSRGGKLEKGLHVGVSVANIGIDIGNSTGGFTDMQLDGAKLVVADVGTNQL 118			
y	y	124 HEKIYTNQEVINIEGVNLRTASKDLIPEVD-----ILTDIVSFISLTULIPSCIRWLKA 178			
b	b	119 VKLKRQDPVRSMEQYNFRVAQ---PEDFNQGQPVFASIDVSF+SLSLILPALHNVLSD 174			
y	y	179 SGIIIALIKPQELPYDDKI-KKGWKETSLOYEAWEKIIHFQSELGGLIFIGVPSVIKG 237			
b	b	175 QGQVIALIKPQFAGEHQIGKGIKVDKQIHKEVIOKVMDFA-SGYGFTVRGLDFSPIQG 233			
y	y	238 PKGNQEYLIVLYKK 250			
b	b	234 GHGNIEFLAHIAK 246			
RESULT 24					
C	D	Q8P077	PRELIMINARY;	PRT;	275 AA.
C	D	Q8P077;			
I	T	01-OCT-2002	(TrEMBLrel. 22, Created)		
I	T	01-OCT-2002	(TrEMBLrel. 22, Last sequence update)		
T	T	01-OCT-2003	(TrEMBLrel. 25, Last annotation update)		
E	N	Putative hemolysin.			

[illegible]

SQ SEQUENCE 275 AA; 30546 MW; EEF702E084B888FFD CRC64;
Query Match 32.5%; Score 413; DB 16; Length 275;
Best Local Similarity 44.2%; Pred. No. 1.7e-29;
Matches 110; Conservative 33; Mismatches 98; Indels 8; Gaps 5;
QY 5 KYRDELVFLQGLAESREQAKRLIMAGKVTLTNNSTTIPRLKPKGHKYPLESICSLIGV 64
DB 3 KERVDVLAQKGLFREQAKRGVWAGLVVNVINGE----RYDKPEKIDEATELKLKGD 58
QY 65 E-RFVSRGAYKLLTALDFFKIDVKSCICLDAGASTGGFTDCLLQHGASKVYAIADVKGQL 123
DB 59 KUKYVSRGGKLEKALKVVISVDGQITLDIGASTGGFTDVLQNGAQLVYADVGTNQL 118
QY 124 HEKLYTNEQVINTGVNLRAS-KDLIPEDVILITDVSFISLTLLPSCIRMLKASGII 192
DB 119 VWKLRQDERVCSMEQYNFRAEYSPDFERGQTFASIDVSPISLNLIPALSKILTDPGGV 178
QY 183 IALIKPOFELYPPDKI-KKGVVVKTSQYEAVEKLIHFQCSQELGLIFGVVPSVIKPKGN 241
DB 179 VALIKPQFEAGREQIGHGIVKDKSVHETVLEITVTPA-TNYGFTVKELDFSPLOQGHG 237
QY 242 QEYLIYLYKK 250
DB 238 IEFLAYLEK 246
RESULT 28
Q7VA50 PRELIMINARY; PRT; 270 AA.
AC Q7VA50; MEDLINE=22810154; PubMed=12917486;
DT 01-OCT-2003 (TReMBLrel. 25, Created)
DT 01-OCT-2003 (TReMBLrel. 25, Last sequence update)
DE Predicted rRNA methylase.
GN PRO1617
OS Prochlorococcus marinus.
OC Bacteria; Cyanobacteria; Prochlorophytes; Prochlorococcaceae;
OC Prochlorococcus.
CX NCBI_TaxID=1219;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SARG / CCMP 1375 / SS120;
RX MEDLINE=22810154; PubMed=12917486;
RA Dufresne A., Salanoubat M., Partensky F., Artiguenave F., Axmann I.M.,
Barbe V., Duprat S., Galperin M.Y., Koonin E.V., Le Gall F.,
Makarova K.S., Ostrowski M., Ozta S., Robert C., Rogozin I.B.,
Scanlan D.J., Tandeau de Marsac N., Weissbach J., Wincker P.,
Wolf Y.I., Hess W.R.;
RT "Genome sequence of the cyanobacterium Prochlorococcus marinus SS120,
a nearly minimal oxypototrophic genome.";
RL Proc. Natl. Acad. Sci. U.S.A. 100:10020-10025 (2003).
DR EMBL; AE017165; AAQ00661.1; -;
KW Methyltransferase; Complete proteome.
SQ SEQUENCE 270 AA; 29644 MW; 5842FB5EDAAA428C4 CRC64;
Query Match 32.3%; Score 411; DB 16; Length 270;
Best Local Similarity 41.9%; Pred. No. 2.6e-29;
Matches 106; Conservative 38; Mismatches 91; Indels 18; Gaps 6;
QY 5 KYRDELVFLQGLAESREQAKRLIMAGKVTLTNNSTTIPRLKPKGHKYPLESICSLIGV 64
DB 4 KNRLDVHLTKGLAPAREQAQKLIRAGKV-----RDVGNILDKPQEQVSKSELEIVQSS 58
QY 65 ERFVSRGAYKLLTALDFFKIDVKSCICLDAGASTGGFTDCLLQHGASKVYAIADVKGQLH 124
DB 59 PRFVSRGGKLEKALKVVISVDGQITLDIGASTGGFTDCLLQNGAQLVYADVGTNQL 118
QY 125 EKLYTNEQVINTGVNLR-----TASKDLIPEEVDILITDVSFISLTLLPSCIRML- 176
DB 119 WTLRNSRVVLRERTNIRYLKHSIDLYGPSDPLP---SLAVADLSFISLRVLPAIKSLIQ 175
QY 177 KASGIIIALIKPOFELYPPDKI-KKGVVVKTSQYEAVEKLIHFQCSQELGLIFGVVPSVI 235

DB 176 KSKQFALLVKPQFEVGPVGVGVVWDAQSHMDALNIIIDFSRSK-DWKYKGAIASPI 234
QY 236 KGPKNQOEYLIYL 248
DB 235 KGPAGNHEYLWLW 247
RESULT 29
Q89RW3 PRELIMINARY; PRT; 245 AA.
ID Q89RW3;
AC Q89RW3; MEDLINE=22484998; PubMed=12597275;
DT 01-JUN-2003 (TReMBLrel. 24, Created)
DT 01-JUN-2003 (TReMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TReMBLrel. 25, Last annotation update)
DE BL12649 protein.
GN BL12649.
OS Bradyrhizobium japonicum.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Bradyrhizobiaceae; Bradyrhizobium.
CX NCBI_TaxID=375;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=USDA 110;
RX MEDLINE=22484998; PubMed=12597275;
RA Kaneko T., Nakamura Y., Sato S., Minamisawa K., Uchiumi T.,
Sasanoto S., Watanabe A., Idesawa K., Iriuchi M., Kawashima K.,
Kohata M., Matsumoto M., Shimo S., Tsuruoka H., Wada T., Yamada M.,
Tabata S.;
RT "Complete genomic sequence of nitrogen-fixing symbiotic bacterium
Bradyrhizobium japonicum USDA110.";
RL DNA Res. 9:189-197 (2002).
DR EMBL; AP005944; BAC47914.1; -;
DR GO; GO:0003723; F:RNA binding; IEA.
DR InterPro; IPR002877; RmMJ_FtsJ.
DR InterPro; IPR002942; S4.
DR Pfam; PF01728; FtsJ; 1.
DR Pfam; PF01479; S4; 1.
DR PROSITE; PS50889; S4; 1.
KW Complete proteome.
SQ SEQUENCE 245 AA; 26173 MW; EC3915FB562C75B5 CRC64;
Query Match 32.1%; Score 408.5; DB 16; Length 245;
Best Local Similarity 38.2%; Pred. No. 3.9e-29;
Matches 94; Conservative 45; Mismatches 100; Indels 7; Gaps 2;
QY 1 MAKHKVRADELVFLQGLAESREQAKRLIMAGKVTLTNNSTTIPRLKPKGHKYPLESICS 60
DB 2 MSPSRKADILLVERGLFESRARARALEAGLVTDKQVT-----KPSETIAEDAVIQ 55
QY 61 LIGVERFVSRGAYKLLTALDFFKIDVKSCICLDAGASTGGFTDCLLQHGASKVYAIADV 120
DB 56 AEPAPHYVSRGGVKGALAGERYPIEDHVCVDVGASTGGFTVLLANGASLVFAIDVGT 115
QY 121 GOLHEKLYTNEQVINTGVNLRASKDLIPEDVILITDVSFISLTLLPSCIRMLKASG 180
DB 116 SOLHPSLRDHPKIVSMEETDIRAVEGRRLPARPDVVVDVFSISKAYLFLVALSIAAPM 175
QY 181 IIALIKPOFELYPPDKIKKGVVVKTSQYEAVEKLIHFQCSQELGLIFGVVPSVIKPKG 240
DB 176 SILLALIKPQFEAGREQIGHGIVKDKSVHETVLEITVTPA-LGCTDIATVFPSPITGG 234
QY 241 NOEYLI 246
DB 235 NIEFFL 240
RESULT 30
Q9K972 PRELIMINARY; PRT; 272 AA.
ID Q9K972;
AC Q9K972; MEDLINE=22484998; PubMed=12597275;
DT 01-OCT-2000 (TReMBLrel. 15, Created)
DT 01-OCT-2000 (TReMBLrel. 15, Last sequence update)

RT emerging human pathogen, serotype V *Streptococcus agalactiae*.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:12391-12396 (2002).
DR EMBL; AE014214; AA099401.1; -.
DR TIGR; SAG0499; -.
DR GO; GO:0003723; F:RNA binding; IEA.
DR GO; GO:0015070; F:toxin activity; IEA.
DR InterPro; IPR002877; RrmJ_FtsJ.
DR InterPro; IPR002942; S4.
DR InterPro; IPR004538; Tly.
DR Pfam; PF01728; FtsJ; 1.
DR Pfam; PF01479; S4; 1.
DR SMART; SM00363; S4; 1.
DR TIGRFAMs; TIGR00478; tly; 1.
DR PROSITE; PS50889; S4; 1.
DR Complete proteome.
SQ SEQUENCE 275 AA; 31014 MW; 83D4FED355247313 CRC64;

Query Match 31.9%; Score 406; DB 16; Length 275;
Best Local Similarity 42.3%; Pred. No. 7.6e-29;
Matches 107; Conservative 44; Mismatches 92; Indels 10; Gaps 6;

QY 1 MAKHKVRADELVFLQGLAESREQAKRLIMAGKVTLTNNSTTIPLRLKPGHKYPLESICS 60
DB 1 MAKERV--DVLAYKQGLFDTRQAKRGVMGNVINNGE----RYDKPGEKVADDTLKL 54

QY 61 LIGVE-RFVSRGAYKLLTALDFKIDVKSICLDAGASTGGTDCLLQHGASKVYVIDVG 119
DB 55 LKGEKLVSRGGKLEKALQVFLSVADKUTIDIGASTGGFTDVMQSGRLVYADVVG 114

QY 120 KGQLEKLYTNEQVINIEGVNLTASKDLIPEEV--DILTIDVSFISLTLLPSCIRWLKA 178
DB 115 TNQLVWLKLRQHRVSRMSQYFNRYAQQEDFEKGLPEPASIDVSFISLTLLPALKEILVD 174

QY 179 SGIITLIKPOFELYPDKI-KGVVVKETSLOYEAVEKLIHFCQSELGLIFIGVPSVILKG 237
DB 175 GGQVVALIKPOFEAGREQIKNGIVKDKLVHEKVLTTVTNFTK-DYGVTVKHLDFSPIQG 233

QY 238 PKGNQEVLYLYLK 250
DB 234 GHGNIFFLMLQK 246

RESULT 33
Q9L3Q7
ID Q9L3Q7 PRELIMINARY; PRT; 281 AA.
AC Q9L3Q7;
DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE Putative haemolysin.
GN TLYA.
OS Mycobacterium ulcerans.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1809;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 19423;
RA Mve Oblang A., Gomez Lopez A., Portaeis F., Fonteyne P.A.;
RT "Cloning and expression of a Mycobacterium ulcerans hemolysin gene.";
RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ271681; CAB83047.1; -.
DR GO; GO:0003723; F:RNA binding; IEA.
DR GO; GO:0015070; F:toxin activity; IEA.
DR InterPro; IPR002877; RrmJ_FtsJ.
DR InterPro; IPR002942; S4.
DR InterPro; IPR004538; Tly.
DR Pfam; PF01728; FtsJ; 1.
DR Pfam; PF01479; S4; 1.
DR SMART; SM00363; S4; 1.
DR TIGRFAMs; TIGR00478; tly; 1.
DR PROSITE; PS50889; S4; 1.
DR Complete proteome.
SQ SEQUENCE 281 AA; 29087 MW; 66FEB33E1156DFE5 CRC64;

RT emerging human pathogen, serotype V *Streptococcus agalactiae*.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:12391-12396 (2002).
DR EMBL; AE014214; AA099401.1; -.
DR TIGR; SAG0499; -.
DR GO; GO:0003723; F:RNA binding; IEA.
DR GO; GO:0015070; F:toxin activity; IEA.
DR InterPro; IPR002877; RrmJ_FtsJ.
DR InterPro; IPR002942; S4.
DR InterPro; IPR004538; Tly.
DR Pfam; PF01728; FtsJ; 1.
DR Pfam; PF01479; S4; 1.
DR SMART; SM00363; S4; 1.
DR TIGRFAMs; TIGR00478; tly; 1.
DR PROSITE; PS50889; S4; 1.
DR Complete proteome.
SQ SEQUENCE 275 AA; 31014 MW; 83D4FED355247313 CRC64;

Query Match 31.9%; Score 406; DB 16; Length 275;
Best Local Similarity 42.3%; Pred. No. 7.6e-29;
Matches 107; Conservative 44; Mismatches 92; Indels 10; Gaps 6;

QY 1 MAKHKVRADELVFLQGLAESREQAKRLIMAGKVTLTNNSTTIPLRLKPGHKYPLESICS 60
DB 1 MAKERV--DVLAYKQGLFDTRQAKRGVMGNVINNGE----RYDKPGEKVADDTLKL 54

QY 61 LIGVE-RFVSRGAYKLLTALDFKIDVKSICLDAGASTGGTDCLLQHGASKVYVIDVG 119
DB 55 LKGEKLVSRGGKLEKALQVFLSVADKUTIDIGASTGGFTDVMQSGRLVYADVVG 114

QY 120 KGQLEKLYTNEQVINIEGVNLTASKDLIPEEV--DILTIDVSFISLTLLPSCIRWLKA 178
DB 115 TNQLVWLKLRQHRVSRMSQYFNRYAQQEDFEKGLPEPASIDVSFISLTLLPALKEILVD 174

QY 179 SGIITLIKPOFELYPDKI-KGVVVKETSLOYEAVEKLIHFCQSELGLIFIGVPSVILKG 237
DB 175 GGQVVALIKPOFEAGREQIKNGIVKDKLVHEKVLTTVTNFTK-DYGVTVKHLDFSPIQG 233

QY 238 PKGNQEVLYLYLK 250
DB 234 GHGNIFFLMLQK 246

RESULT 33
Q9L3Q7
ID Q9L3Q7 PRELIMINARY; PRT; 281 AA.
AC Q9L3Q7;
DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE Putative haemolysin.
GN TLYA.
OS Mycobacterium ulcerans.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1809;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 19423;
RA Mve Oblang A., Gomez Lopez A., Portaeis F., Fonteyne P.A.;
RT "Cloning and expression of a Mycobacterium ulcerans hemolysin gene.";
RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ271681; CAB83047.1; -.
DR GO; GO:0003723; F:RNA binding; IEA.
DR GO; GO:0015070; F:toxin activity; IEA.
DR InterPro; IPR002877; RrmJ_FtsJ.
DR InterPro; IPR002942; S4.
DR InterPro; IPR004538; Tly.
DR Pfam; PF01728; FtsJ; 1.
DR Pfam; PF01479; S4; 1.
DR SMART; SM00363; S4; 1.
DR TIGRFAMs; TIGR00478; tly; 1.
DR PROSITE; PS50889; S4; 1.
DR Complete proteome.
SQ SEQUENCE 281 AA; 29087 MW; 66FEB33E1156DFE5 CRC64;

Query Match 31.5%; Score 400; DB 16; Length 266;
Best Local Similarity 37.5%; Pred. No. 2.6e-28;
Matches 95; Conservative 51; Mismatches 95; Indels 12; Gaps 4;

QY 3 KHKVRADELVFLQGLAESREQAKRLIMAGKVTLTNNSTTIPLRLKPGHKYPLESICS 62
DB 6 KNNKRLDEYLVENKYPENLEIAKQIMVGNVIVNER-----KIDKPGELIILDKVKSIR 59

Query Match 31.7%; Score 403.5; DB 2; Length 281;
Best Local Similarity 42.1%; Pred. No. 1.3e-28;
Matches 106; Conservative 40; Mismatches 95; Indels 11; Gaps 6;

QY 1 MAKHKVRADELVFLQGLAESREQAKRLIMAGKVTLTNNSTTIPLRLKPGHKYPLESICS 60
DB 1 MAR--RTRVDAELVRRGLARSQQAELISAGKVRIDG-----LPAYKPATAVAATAALT 53

QY 61 LIG-VER-FVSRGAYKLLTALDFKIDVKSICLDAGASTGGTDCLLQHGASKVYVIDV 118
DB 54 VAADSERGWSRGAKHLIGALDAFGIDVHGRCCLDAGASTGGFTVLLDRGAEEVAAADV 113

QY 119 KGQLEKLYTNEQVINIEGVNLTASKDLIPEEV--DILTIDVSFISLTLLPSCIRWLKA 178
DB 114 GYGQLAWSLATDPRVIVERTNVRDLSPAATGGQVLDLVADLSFISLTTLVLAALACALP 173

QY 179 SGIITLIKPOFELYPDKI-KGVVVKETSLOYEAVEKLIHFCQSELGLIFIGVPSVILKG 237
DB 174 HADIVPMVKPQFEVKGQGVGAGGVVQDPALRAGAVLSVARRA-GELGWHTVGVTTASPLPG 232

QY 238 PKGNQEVLYLYLK 249
DB 233 PSQNVFELMLDR 244

RESULT 34
Q8REAS
ID Q8REAS PRELIMINARY; PRT; 266 AA.
AC Q8REAS;
DT 01-JUN-2002 (TREMBlrel. 21, Created)
DT 01-JUN-2002 (TREMBlrel. 21, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE Hemolysin.
GN FN1206.
OS Fusobacterium nucleatum (subsp. nucleatum).
OC Bacteria; Fusobacteria; Fusobacteriales; Fusobacteriaceae;
OC Fusobacterium.
OX NCBI_TaxID=76856;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 25586;
RX MEDLINE=21886394; PubMed=11889109;
RA Kapatal V., Anderson I., Ivanova N., Reznik G., Los T., Lykidis A.,
RA Bhattacharya A., Bartman A., Gardner W., Grechkin G., Zhu L.,
RA Vasieva O., Chu L., Kogan Y., Chaga O., Goltzman E., Bernal A.,
RA Larsen N., D'Souza M., Walunas T., Pusch G., Haselkorn R.,
RA Foust M., Kyrpides N., Overbeek R.;
RT "Genome sequence and analysis of the oral bacterium Fusobacterium
nucleatum strain ATCC 25586.";
RL J. Bacteriol. 184:2005-2018(2002).
DR EMBL; AE010627; AAL95402.1; -.
DR GO; GO:0003723; F:RNA binding; IEA.
DR GO; GO:0015070; F:toxin activity; IEA.
DR InterPro; IPR002877; RrmJ_FtsJ.
DR InterPro; IPR002942; S4.
DR InterPro; IPR004538; Tly.
DR Pfam; PF01728; FtsJ; 1.
DR Pfam; PF01479; S4; 1.
DR SMART; SM00363; S4; 1.
DR TIGRFAMs; TIGR00478; tly; 1.
DR PROSITE; PS50889; S4; 1.
DR Complete proteome.
SQ SEQUENCE 266 AA; 30738 MW; 7BD8BE2466861EDF CRC64;

Query Match 31.5%; Score 400; DB 16; Length 266;
Best Local Similarity 37.5%; Pred. No. 2.6e-28;
Matches 95; Conservative 51; Mismatches 95; Indels 12; Gaps 4;

QY 3 KHKVRADELVFLQGLAESREQAKRLIMAGKVTLTNNSTTIPLRLKPGHKYPLESICS 62
DB 6 KNNKRLDEYLVENKYPENLEIAKQIMVGNVIVNER-----KIDKPGELIILDKVKSIR 59

Qy	63	GVER---	FVSGAYKLLTALDFFKIDVKVSCICLDAGASTGGFTDCLQHGASKVYADV	119
Db				
Db	60	IKEDSPVSRGGLKEAKIVFDLDFDKIILDIGASTGGFTDCSLQNGAKFVYADV	119	
Qy	120	KGQHEKLYTNEQVINTGCVNLRASKDLIPEVDILTDVVSFISLTLILPSCIRWLKAS	179	
Db				
Db	120	TNQJDKWLNRDCRVKSIENHNLEKSLDKDDIDIIVMDISFISKVLYIKIRLSKEN	179	
Qy	180	GIITAIKPPQPELYPDKIKGWKVBTSLOYBAVEKIIHFQCSGELGIFT-GVVPVSIKGP	238	
Db				
Db	180	GFAIFLIKPPQEAENEIEKGVDDPNVHKRVINEVB--EAKIHQLFLENLTVPFIKGT	237	
Qy	239	KGQOEYLIYLKKR	251	
Db				
Db	238	KGNIETYLAKFSKX	250	

RESULT 35
Q9F6Y3
ID Q9F6Y3 PRELIMINARY; PRT; 260 AA.
AC Q9F6Y3;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE HlyA.
GN HLYA.
OS Chloroflexus aurantiacus.
OC Bacteria; Chloroflexi; Chloroflexales; Chloroflexaceae; Chloroflexus.

RESULT 36	Q92RR	Hypot
Q92RJ2	Q92RR	R008
ID	AC	Rhiz
DT	DT	Bact
DT	DT	Rhiz
DT	DT	NCBI
DT	DT	[1]
DE	DE	SEQU
GN	GN	STR
OS	OS	
OC	OC	
OC	OC	
OX	OX	
RN	RN	
RP	RP	
RC	RC	

```

RESULT 37
ID 005676
ID 005676 PRELIMINARY; PRT: 269 AA.
AC 005676;
DT 01-JUL-1997 (TrEMBLrel. 04, Created)
DT 01-JUL-1997 (TrEMBLrel. 04, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein (cytotoxin/hemolysin).
GN TLJA OR ML1358 OR MLC1351.14C.

```

OS Mycobacterium leprae.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1769;
[1]
RN SEQUENCE FROM N.A.
RA Murphy L., Harris D.;
RL Submitted (MAY-1997) to the EMBL/GenBank/DBJ databases.
[2]
RN SEQUENCE FROM N.A.
RC STRAIN=TN;
RX MEDLINE=21128732; PubMed=11234002;
RA Cole S.T., Eiglmeier K., Parkhill J., James K.D., Thomson N.R.,
RA Wheeler P.R., Honore N., Garnier T., Churcher C., Harris D.,
RA Mungall K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R.M., Devlin K., Duthoy S., Felwell T., Fraser A., Hamlin N.,
RA Holroyd S., Hornsby T., Jagels K., Lacroix C., Maclean J., Moule S.,
RA Murphy L., Oliver K., Quail M.A., Rajandream M.A., Rutherford K.M.,
RA Rutter S., Seeger K., Simon S., Simmonds M., Skelton J., Squares R.,
RA Squares S., Stevens K., Taylor K., Whitehead S., Woodward J.R.,
RA Barrell B.G.;
RL "Massive gene decay in the leprosy bacillus.";
RT Nature 409:1007-1011(2001).
DR EMBL; Z95117; CAB08287.1; -.
DR EMBL; AL583921; CAC31739.1; -.
DR PIR; H87078; H87078.
DR Leproma; MLI358; -.
DR GO; GO:0003723; F:RNA binding; IEA.
DR GO; GO:0015070; F:toxin activity; IEA.
DR InterPro; IPR002877; RmmJ_FtsJ.
DR InterPro; IPR002942; S4.
DR InterPro; IPR004538; Tly.
DR Pfam; PF01728; FtsJ; 1.
DR SMART; SM00363; S4; 1.
DR TIGR; TIGR00478; tly; 1.
DR PROSITE; PS50889; S4; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 269 AA; 28203 MW; 739B78D8AF5EA043 CRC64;

Query Match 30.3%; Score 384.5; DB 16; Length 269;
Best Local Similarity 38.9%; Pred. No. 6.8e-27;
Matches 98; Conservative 49; Mismatches 94; Indels 11; Gaps 5;

Qy 1 MAKHKVRADELVFLQGLAESREQAKRLIMAGKVTILTNNSITPILEKPGHKYPLESICS 60
Db 1 MAR-RVVDVDELVRGLARSRQQAELIGAGKVRIDG-----LPVAKPATAVSTTALT 53
Qy 61 LI--GVERFVSGYKLLTALDFFKIDVKSCICLDAGASTGGFTDCLLQHGASKYVAIDV 118
Db 54 VADGERSWVSRGAHKLMGALDTFGIPVAGRCCLDAGASTGGFTFVLLDRGAHVAVDV 113
Qy 119 GKGLHEKLYTNEQVINTGVNLTASKDLIPEEVDIITDVFSFSLTLILPSCIRWLKA 178
Db 114 GYGQLAWSVRCDPRVIVVETNTVHDLSPELIGPVDLVVADLSFSLATVLSALAGCALP 173
Qy 179 SGIIIALIKPQELYPDKI-KGVVKTSLQVEAVEKIIHFQCSGLHIFGVPSVTKG 237
Db 174 SADIIVPMVKPQFVGVGQGVGVVRLRLRLRADSVLAVARRA-TELGWRTMDVTASSLPG 232
Qy 238 PKGNOEYLILYK 249
Db 233 LSGNVEYFLWLR 244

RESULT 38

Q50760 PRELIMINARY; PRT; 268 AA.
AC Q50760; 007264;
DT 01-NOV-1996 (TReMBLrel. 01, Created)
DT 01-JAN-1998 (TReMBLrel. 05, Last sequence update)
DT 01-OCT-2003 (TReMBLrel. 25, Last annotation update)
DE TLXA protein (Cytotoxin/hemolysin).

GN TLYA OR RV1694 OR MTC125.16 OR MT1733.
OS Mycobacterium tuberculosis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
[1]
RN SEQUENCE FROM N.A.
RA Wren B., Stabler R.;
RL Submitted (AUG-1997) to the EMBL/GenBank/DBJ databases.
[2]
RN SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Felwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RL "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence.";
RN Nature 393:537-544(1998).
[3]
RN SEQUENCE FROM N.A.
RA Wren B.W., Stabler R.A., Das S.S., Butcher P., Monahan I., Casali N.,
RA Parish T., Stoker N.G.;
RL Submitted (JUN-1996) to the EMBL/GenBank/DBJ databases.
[4]
RN SEQUENCE FROM N.A.
RC STRAIN=CDC 1551 / Oshkosh;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA Bishai W.;
RL "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT laboratory strains";
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; X98295; CAA66941.1; -.
DR EMBL; Z98268; CAB10951.1; -.
DR EMBL; AE007035; AAK46002.1; -.
DR PIR; E70502; E70502.
DR TIGR; MT1733; -.
DR TubercuList; RV1694; -.
DR GO; GO:0003723; F:RNA binding; IEA.
DR GO; GO:0015070; F:toxin activity; IEA.
DR InterPro; IPR002877; RmmJ_FtsJ.
DR InterPro; IPR002942; S4.
DR InterPro; IPR004538; Tly.
DR Pfam; PF01728; FtsJ; 1.
DR SMART; SM00363; S4; 1.
DR TIGR; TIGR00478; tly; 1.
DR PROSITE; PS50889; S4; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 268 AA; 28074 MW; B62AB8400CCC71CE CRC64;

Query Match 29.6%; Score 376.5; DB 16; Length 268;
Best Local Similarity 38.9%; Pred. No. 3.7e-26;
Matches 98; Conservative 42; Mismatches 101; Indels 11; Gaps 5;

Qy 1 MAKHKVRADELVFLQGLAESREQAKRLIMAGKVTILTNNSITPILEKPGHKYPLESICS 60
Db 1 MAR-RARVDAELVRGLARSRQQAELIGAGKVRIDG-----LPVAKPATAVSTTALT 53
Qy 61 LI--GVERFVSGYKLLTALDFFKIDVKSCICLDAGASTGGFTDCLLQHGASKYVAIDV 118
Db 54 VVTSERAWVSRGAHKLVGALFAIAVAGRCCLDAGASTGGFTFVLLDRGAHVAVDV 113
Qy 119 GKGLHEKLYTNEQVINTGVNLTASKDLIPEEVDIITDVFSFSLTLILPSCIRWLKA 178

Db 114 GYGQAWSLNDPRVVLRTNARGLTPEAIGRVDLVADLSTFISLATVLPALVGCASR 173
QY 179 SGIIIALIKQFELYPKI-KKGVVKETSLOYEAVEKIIHFCQSELGLIFIGVPSVIK 237
Db 174 DADIVLVKQFEGVGQGGVGHDPQLRARSVLAVARRAQ-ELGWHSGVGKASPLPG 232
QY 238 PKGNQRYLYLK 249
Db 233 PSGNVEYFLWLR 244

RESULT 39

Q7VEU4 PRELIMINARY; PRT; 268 AA.
ID Q7VEU4
AC Q7VEU4
DT 01-OCT-2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DE 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE CYTOTOXIN|HEMOLYSIN HOMOLOGUE TLXA.
GN TLXA OR MB1720.
OS Mycobacterium bovis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1765;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=AF2122/97;
RX MEDLINE=22709107; PubMed=12788972;
RA Garnier T., Eiglmeter K., Camus J.-C., Medina N., Mansoor H.,
RA Pyvor M., Duthoy S., Grondin S., Lacroix C., Monsempe C., Simon S.,
RA Harris B., Atkin R., Doggett J., Mayes R., Keating L., Wheeler P.R.,
RA Parkhill J., Barrall B.G., Cole S.T., Gordon S.V., Hewinson R.G.;
RL "The complete genome sequence of Mycobacterium bovis";
RT Proc. Natl. Acad. Sci. U.S.A. 100:7877-7882(2003).
DR EMBL; BX248339; CAD96388.1; --
KW Complete proteome.
SQ SEQUENCE 268 AA; B62AB8400CCC71CE CRC64;

Query Match 29.6%; Score 376.5; DB 16; Length 268;
Best Local Similarity 38.9%; Pred. No. 3.7e-26;
Matches 98; Conservative 49; Mismatches 101; Indels 11; Gaps 5;
QY 1 MAKHKVRADELVLQGLAESREQAKRLIMAGKVTLTNNSTIPLRLKPGHKYPLESICS 60
Db 1 MAR-RARVDLVRGLARSQQAELIGAGKVRIDG-----LPVAKPATVSDTTALT 53
QY 61 LI--GVERFVSRGAYKLLTALDFFKIDVKSCICLDAGASTGGTDCILQHGASKVAIDV 118
Db 54 VVTDSEAWVSRCAGKLVGLALEFAFATAVAGRCILDAGASTGGTTEVLLDRGAHVVAADV 113
QY 119 KGQOLHEKLYTNEQVNIENGVNLTASKDLIPEEVDILTIDVFSILTLPSCIRWLKA 178
Db 114 GYGQAWSLNDPRVVLRTNARGLTPEAIGRVDLVADLSTFISLATVLPALVGCASR 173
QY 179 SGIIIALIKQFELYPKI-KKGVVKETSLOYEAVEKIIHFCQSELGLIFIGVPSVIK 237
Db 174 DADIVLVKQFEGVGQGGVGHDPQLRARSVLAVARRAQ-ELGWHSGVGKASPLPG 232
QY 238 PKGNQRYLYLK 249
Db 233 PSGNVEYFLWLR 244

RESULT 40

Q9LSV5 PRELIMINARY; PRT; 301 AA.
ID Q9LSV5
AC Q9LSV5
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DE 01-OCT-2000 (TrEMBLrel. 25, Last annotation update)
DE Hemolysin-like (Hemolysin, putative).
OS Arabidopsis thaliana (Mouse-ear cross).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Columbia;
RA Sato S., Nakamura Y., Kaneko T., Kato T., Asamizu E., Tabata S.;
RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Columbia;
RX MEDLINE=20277480; PubMed=10819329;
RA Nakamura Y.;
RT "Structural analysis of Arabidopsis thaliana chromosome 3. I. Sequence
RT features of the regions of 4,504,864 bp covered by sixty P1 and TAC
RL clones.";
RN DNA Res. 7:131-135(2000).
RN [3]
RP SEQUENCE FROM N.A.
RA Haas B.J., Volkovskiy N., Town C.D., Troukhan M., Alexandrov N.,
RA Feldmann K.A., Flavell R.B., White O., Salzberg S.L.;
RT "Full-length messenger RNA sequences greatly improve genome
RT annotation.";
RL Genome Biol. 0:0-0(2002).
RN [4]
RP SEQUENCE FROM N.A.
RA Brover V., Troukhan M., Alexandrov N., Lu Y.-P., Flavell R.,
RA Feldmann K.;
RT "Full-length cDNA from Arabidopsis thaliana";
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB025639; BAB01316.1; --
DR EMBL; AY087760; BAM65296.1; --
DR GO; GO:0003723; F:RNA binding; IEA.
DR GO; GO:0008757; F:S-adenosylmethionine-dependent methyltransf. .; IEA.
DR GO; GO:0015070; F:toxin activity; IEA.
DR InterPro; IPR002877; RsmJ_FtsJ.
DR InterPro; IPR002942; S4.
DR InterPro; IPR000051; SAM_bind.
DR InterPro; IPR004538; Tly.
DR Pfam; PF01728; FtsJ; 1.
DR Pfam; PF01479; S4; 1.
DR TIGRFAMs; TIGR00478; tly; 1.
DR PROSITE; PS00889; S4; 1.
SQ SEQUENCE 301 AA; 33204 MW; ACCCIF0E01B114A40 CRC64;

Query Match 29.5%; Score 374.5; DB 10; Length 301;
Best Local Similarity 37.1%; Pred. No. 6.5e-26;
Matches 91; Conservative 49; Mismatches 90; Indels 15; Gaps 6;
QY 5 KVRADLVLFLQGLAESREQAKRLIMAGKVTLTNNSTIPLRLKPGHKYPLESICS--- 61
Db 54 KQRLDEACVFERFNEYSTRLIQSWIMQGVLDGR-----RVIKAG--MPVSTAAIKIT 105
QY 62 IGVRFVSRGAYKLLTALDFFKIDVKSCICLDAGASTGGTDCILQHGASKVAIDVGK 121
Db 106 AEVPKFCVCGGLKLEALEKLDVDVSEKVLVDAGLSTGGTDCILRYGAHVGVGVG 165
QY 122 QLHEKLYTNEQVNIENGVNLTASKDLIPEEVDILTIDVFSILTLPSCIRWLKASGI 181
Db 166 QVADKTRNDKVTIVERTNRLVLPG--LPQKVDVTLDSFISILKVMALMVMKEDAT 223
QY 182 IIALIKQFELYPKI-KKGVVKETSLOYEAVEKIIHFCQSELGLIFIGVPSVIKPGK 240
Db 224 LVTLVKQFARRQQVGKGIIVRDPVHQEVLEKIINGVE-RYGFTNKGFIESPICKGADG 282
QY 241 NOEYL 245
Db 283 NIEFL 287

Search completed: July 7, 2004, 16:56:43
Job time : 66.3721 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 7, 2004, 16:52:58 ; Search time 5.48173 Seconds
(without alignments)
470.891 Million cell updates/sec

Title: US-10-009-919a-1_COPY_1_50

Perfect score: 245

Sequence: 1 MAKHKVRADELVFLOGLAES.....GKVTLTNNSTTPIRLKPG 50

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- Issued Patents AA.*
- 1: /cgn2_6/ptodata/2/iaa/5A COMB.pep.*
 - 2: /cgn2_6/ptodata/2/iaa/5B COMB.pep.*
 - 3: /cgn2_6/ptodata/2/iaa/6A COMB.pep.*
 - 4: /cgn2_6/ptodata/2/iaa/6B COMB.pep.*
 - 5: /cgn2_6/ptodata/2/iaa/PTUS COMB.pep.*
 - 6: /cgn2_6/ptodata/2/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	106.5	43.5	276	4	US-09-634-238-259
2	95	38.8	270	4	US-08-961-083-64
3	95	38.8	270	4	US-09-536-784-64
4	90.5	36.9	227	4	US-09-107-532A-6779
5	74	30.2	137	4	US-08-936-165A-431
6	73	29.8	220	4	US-09-134-001C-5073
7	61	24.9	240	1	US-07-965-668A-3
8	61	24.9	240	2	US-08-950-433-3
9	61	24.9	240	3	US-09-186-287-3
10	59	24.1	160	4	US-09-732-210-1604
11	58	23.7	207	4	US-09-252-991A-21381
12	57	23.3	357	3	US-08-960-780-38
13	57	23.3	357	3	US-09-073-898-38
14	57	23.3	357	3	US-09-850-351A-38
15	56.5	23.1	220	4	US-09-134-000C-3991
16	56.5	23.1	476	4	US-09-489-039A-8358
17	55	22.4	411	4	US-09-252-991A-24632
18	54	22.0	153	4	US-09-540-236-3643
19	54	22.0	479	4	US-09-723-368-2
20	53.5	21.8	480	4	US-09-252-991A-19202
21	53	21.6	171	4	US-09-732-210-77
22	53	21.6	212	4	US-09-328-352-4325
23	53	21.6	252	4	US-09-199-637A-209
24	53	21.6	330	4	US-09-252-991A-29761
25	53	21.6	427	4	US-09-199-637A-107
26	52	21.2	117	4	US-09-107-532A-6325
27	51.5	21.0	239	4	US-09-107-532A-5582

28 51.5 21.0 409 4 US-09-328-352-6650 Sequence 6650, Ap
29 51 20.8 285 4 US-09-134-000C-4654 Sequence 4654, Ap
30 51 20.8 376 4 US-09-328-352-7587 Sequence 7587, Ap
31 51 20.8 463 4 US-09-252-991A-24757 Sequence 24757, A
32 51 20.8 566 4 US-09-513-783A-142 Sequence 142, App
33 51 20.8 812 4 US-09-513-783A-4 Sequence 4, Appli
34 51 20.8 812 4 US-09-513-783A-6 Sequence 6, Appli
35 51 20.8 1056 4 US-09-513-783A-32 Sequence 32, Appl
36 51 20.8 1125 4 US-09-513-783A-152 Sequence 152, App
37 51 20.8 1610 4 US-09-513-783A-22 Sequence 22, Appl
38 50 20.4 240 4 US-09-107-532A-6634 Sequence 6634, Ap
39 50 20.4 315 4 US-09-134-000C-5324 Sequence 5324, Ap
40 50 20.4 316 4 US-09-107-532A-4924 Sequence 4924, Ap
41 50 20.4 421 4 US-10-067-443-4 Sequence 4, Appli
42 50 20.4 421 4 US-10-067-443-28 Sequence 28, Appl
43 50 20.4 463 4 US-09-489-039A-12834 Sequence 12834, A
44 50 20.4 633 4 US-09-976-594-282 Sequence 282, App
45 49.5 20.2 313 3 US-09-413-814-9 Sequence 9, Appli

ALIGNMENTS

RESULT 1

US-09-634-238-259
; Sequence 259, Application US/09634238
; Patent No. 6544772

GENERAL INFORMATION:
; APPLICANT: Glenn, Matthew

; APPLICANT: Havukkala, Ilkka J.
; APPLICANT: Bloksberg, Leonard, N.

; APPLICANT: Lubbers, Mark W.

; APPLICANT: Dekker, James

; APPLICANT: Christensson, Anna C.

; APPLICANT: Holland, Ross

; APPLICANT: O'Toole, Paul W.

; APPLICANT: Reid, Julian R.

; APPLICANT: Coolbear, Timothy

; TITLE OF INVENTION: Polynucleotides, materials incorporating them and methods for using them.

; FILE REFERENCE: 11000.1043U1

; CURRENT APPLICATION NUMBER: US/09/634,238

; CURRENT FILING DATE: 2000-08-08

; NUMBER OF SEQ ID NOS: 422

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 259

; LENGTH: 276

; TYPE: PRT

; ORGANISM: Lactobacillus rhamnosus

US-09-634-238-259

Query Match 43.5%; Score 106.5; DB 4; Length 276;

Best Local Similarity 56.5%; Pred. No. 1.1e-07;

Matches 26; Conservative 4; Mismatches 11; Indels 5; Gaps 1;

QY 5 KVRADLVFLOGLAESREQAKRLIMAGKVTLTNNSTTPIRLKPG 50

Db 6 KERVDVLLVEQGLFESREQAKRSVMAGEVYDQNNQ-----RLDKPG 46

RESULT 2

US-08-961-083-64

; Sequence 64, Application US/08961083

; Patent No. 6159469

GENERAL INFORMATION:

; APPLICANT: Choi et. al.

; TITLE OF INVENTION: Streptococcus pneumoniae Antigenes and Vaccines

; NUMBER OF SEQUENCES: 452

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Human Genome Sciences, Inc.

; STREET: 9410 Key West Avenue

; CITY: Rockville

; STATE: Maryland

Mon Jul 12 15:30:39 2004

```
;
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/961,083
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Brookes, A. Anders
; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: PB340P2
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 64:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 270 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-961-083-64
;
; Query Match 38.8%; Score 95; DB 3; Length 270;
; Best Local Similarity 50.0%; Pred. No. 5.3e-06;
; Matches 23; Conservative 4; Mismatches 15; Indels 4; Gaps 1;
;
; QY 5 KVRADLVFLQGLAESEQAKRLIMAGKVTLTNNSTIPLRLKPG 50
; | | | | | : | | | | | : | | | | | : | | | | |
; Db 2 KERVVDVLAYKQGLFETREQAKRGVMAGLVVAVLNGE----RFDKPG 43
;
; RESULT 3
; US-09-536-784-64
; Sequence 64, Application US/09536784
; Patent No. 6573082
; GENERAL INFORMATION:
; APPLICANT: Choi et. al.
; TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines
; NUMBER OF SEQUENCES: 452
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/536,784
; FILING DATE: 30-Oct-1997
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/961,083
; FILING DATE: OCT-30-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Michelle S. Marks
; REGISTRATION NUMBER: 41,971
; REFERENCE/DOCKET NUMBER: PB340P3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
```

```
;
; INFORMATION FOR SEQ ID NO: 64:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 270 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 64:
; US-09-536-784-64
;
; Query Match 38.8%; Score 95; DB 4; Length 270;
; Best Local Similarity 50.0%; Pred. No. 5.3e-06;
; Matches 23; Conservative 4; Mismatches 15; Indels 4; Gaps 1;
;
; QY 5 KVRADLVFLQGLAESEQAKRLIMAGKVTLTNNSTIPLRLKPG 50
; | | | | | : | | | | | : | | | | | : | | | | |
; Db 2 KERVVDVLAYKQGLFETREQAKRGVMAGLVVAVLNGE----RFDKPG 43
;
; RESULT 4
; US-09-107-532A-6779
; Sequence 6779, Application US/09107532A
; Patent No. 6583275
; GENERAL INFORMATION:
; APPLICANT: Lynn A. Doucette-Stamm and David Bush
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS
; NUMBER OF SEQUENCES: 7310
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: GENOME THERAPEUTICS CORPORATION
; STREET: 100 Beaver Street
; CITY: Waltham
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02354
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD-ROM ISO9660
; OPERATING SYSTEM: <Unknown>
; SOFTWARE: ASCII
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/107,532A
; FILING DATE: 30-Jun-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/085,598
; FILING DATE: 14 May 1998
; APPLICATION NUMBER: 60/051571
; FILING DATE: July 2, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Ariniello, Pamela Deneke
; REGISTRATION NUMBER: 40,489
; REFERENCE/DOCKET NUMBER: GTC-012
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (781) 893-5007
; TELEFAX: (781) 893-8277
; INFORMATION FOR SEQ ID NO: 6779:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 227 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: YES
; ORIGINAL SOURCE:
; ORGANISM: Enterococcus faecium
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (B) LOCATION 1...227
; SEQUENCE DESCRIPTION: SEQ ID NO: 6779:
; US-09-107-532A-6779
;
; Query Match 36.9%; Score 90.5; DB 4; Length 227;
; Best Local Similarity 47.8%; Pred. No. 2e-05;
; Matches 22; Conservative 4; Mismatches 15; Indels 5; Gaps 1;
```


QY 5 KRADELVFLQGLAESRQAKRLIMAGKVTLTNNSTTIPRLKPG 50
| | | | | : | | | | | : | | | : | | |
Db 10 KERVDVLAVKQGLFETREQAKRSWAGLIYNEKNE-----RFDKPG 50

RESULT 5
US-08-936-165A-431
; Sequence 431, Application US/08936165A
; Patent No. 6348582
; GENERAL INFORMATION:
; APPLICANT: Black, Michael
; APPLICANT: Burnham, Martin
; APPLICANT: Hodgson, John
; APPLICANT: Knowles, David
; APPLICANT: Lonetto, Michael
; APPLICANT: Nicholas, Richard
; APPLICANT: Pratt, Julie
; APPLICANT: Reichard, Richard
; APPLICANT: Rosenberg, Martin
; APPLICANT: Ward, Judith
; TITLE OF INVENTION: No. 6348582el Prokaryotic Polynucleotides,
; TITLE OF INVENTION: Polypeptides and Their Uses
; NUMBER OF SEQUENCES: 534
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Smithkline Beecham Corporation
; STREET: 709 Swedeland Road
; CITY: King of Prussia
; STATE: PA
; COUNTRY: USA
; ZIP: 19406-0939
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/936,165A
; FILING DATE: 24-SEP-1997
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/027,032
; FILING DATE: 24-SEP-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Gimmi, Edward R
; REGISTRATION NUMBER: 38,891
; REFERENCE/DOCKET NUMBER: P50549
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 610-270-4478
; TELEFAX: 610-270-5090
; TELEX:
; INFORMATION FOR SEQ ID NO: 431:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 137 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: Protein
US-08-936-165A-431

Query Match 30.2%; Score 74; DB 4; Length 137;
Best Local Similarity 36.4%; Pred. No. 0.0029;
Matches 16; Conservative 8; Mismatches 20; Indels 0; Gaps 0;

QY 7 RADELVFLQGLAESRQAKRLIMAGKVTLTNNSTTIPRLKPG 50
| | | | | : | | | | | : | | | : | | |
Db 14 RLDVVYSLGLARTRQARQLVNHGHIIVDGRVDIPSYGVKPG 57

RESULT 6
US-09-134-001C-5073
; Sequence 5073, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 5073
; LENGTH: 220
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-5073

Query Match 24.9%; Score 61; DB 1; Length 240;
Best Local Similarity 38.9%; Pred. No. 0.52;
Matches 14; Conservative 7; Mismatches 15; Indels 0; Gaps 0;

Query Match 29.8%; Score 73; DB 4; Length 220;
Best Local Similarity 36.4%; Pred. No. 0.0076;
Matches 16; Conservative 8; Mismatches 20; Indels 0; Gaps 0;
QY 7 RADELVFLQGLAESRQAKRLIMAGKVTLTNNSTTIPRLKPG 50
| | | | | : | | | | | : | | | : | | |
Db 113 RLDVVYSLGLARTRQARQLVNHGHIIVDGRVDIPSYSLKPG 156

RESULT 7
US-07-965-668A-3
; Sequence 3, Application US/07965668A
; Patent No. 5364774
; GENERAL INFORMATION:
; APPLICANT: MUIR, SUSIE J.
; APPLICANT: KOOPMAN, MARCEL B.H.
; APPLICANT: KUSTERS, JOHANNES G.
; TITLE OF INVENTION: TREPONEMA HYODYSENTERIAE VACCINE
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: STEVENS, DAVIS, MILLER & MOSHER
; STREET: 515 NORTH WASHINGTON STREET
; CITY: ALEXANDRIA
; STATE: VIRGINIA
; COUNTRY: USA
; ZIP: 22314
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/965,668A
; FILING DATE: 19921021
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: POULOS III, JAMES A.
; REGISTRATION NUMBER: 31,714
; REFERENCE/DOCKET NUMBER: EHM 27571
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 703 549 7200
; TELEFAX: 703 528 5313
; TELEX: 44-0704
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 240 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-07-965-668A-3

Query Match 24.9%; Score 61; DB 1; Length 240;
Best Local Similarity 38.9%; Pred. No. 0.52;
Matches 14; Conservative 7; Mismatches 15; Indels 0; Gaps 0;


```
US-09-252-991A-21381
; Sequence 21381, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 21381
; LENGTH: 207
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-21381

Query Match      23.7%; Score 58; DB 4; Length 207;
Best Local Similarity 25.0%; Pred. No. 1.2;
Matches 11; Conservative 11; Mismatches 22; Indels 0; Gaps 0;

QY      7 RADELVFLQGLAESREQAKRLIMAGKVTLTNNSTTPIRLKPG 50
Db      98 RLDNVVVMGFGSTRSSRQLVSHKAITVNGQTWNIPSYQVKAG 141

RESULT 12
US-08-960-780-38
; Sequence 38, Application US/08960780
; Patent No. 6204435
; GENERAL INFORMATION:
; APPLICANT: Feitelson, Jerald S.
; APPLICANT: Schnepf, H. Ernest
; APPLICANT: Narva, Kenneth E.
; APPLICANT: Stockhoff, Brian A.
; APPLICANT: Schmeits, James
; APPLICANT: Loewer, David
; APPLICANT: Dullum, Charles Joseph
; APPLICANT: Muller-Cohn, Judy
; APPLICANT: Stamp, Lisa
; TITLE OF INVENTION: No. 6204435el Pesticidal Toxins and Nucleotide
; TITLE OF INVENTION: Sequences Which Encode These Toxins
; NUMBER OF SEQUENCES: 134
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Saliwanchik, Lloyd & Saliwanchik
; STREET: 2421 N.W. 41st Street, Suite A-1
; CITY: Gainesville
; STATE: FL
; COUNTRY: US
; ZIP: 32606-6669
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/960,780
; FILING DATE: 30-OCT-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/029,848
; FILING DATE: 30-OCT-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Saliwanchik, David R.
; REGISTRATION NUMBER: 31,794
; REFERENCE/DOCKET NUMBER: MA-708
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 352-375-8100
; TELEFAX: 352-372-5800
```

```
; INFORMATION FOR SEQ ID NO: 38:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 357 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; ORIGINAL SOURCE:
; INDIVIDUAL ISOLATE: 196F3
US-08-960-780-38

Query Match      23.3%; Score 57; DB 3; Length 357;
Best Local Similarity 40.6%; Pred. No. 3.5;
Matches 13; Conservative 6; Mismatches 13; Indels 0; Gaps 0;

QY      17 LAESREQAKRLIMAGKVTLTNNSTTPIRLK 48
Db      63 LSTSSDQOVFTLXGKIILNQSSMTPIRLK 94

RESULT 13
US-09-073-898-38
; Sequence 38, Application US/09073898
; Patent No. 6242669
; GENERAL INFORMATION:
; APPLICANT: Feitelson, Jerald S.
; APPLICANT: Schnepf, H. Ernest
; APPLICANT: Narva, Kenneth E.
; APPLICANT: Stockhoff, Brian A.
; APPLICANT: Schmeits, James
; APPLICANT: Loewer, David
; APPLICANT: Dullum, Charles Joseph
; APPLICANT: Muller-Cohn, Judy
; APPLICANT: Stamp, Lisa
; APPLICANT: Morrill, George
; APPLICANT: Finstad-Lee, Stacey
; TITLE OF INVENTION: No. 6242669el Pesticidal Toxins and Nucleotide
; TITLE OF INVENTION: Sequences Which Encode These Toxins
; NUMBER OF SEQUENCES: 144
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Saliwanchik, Lloyd & Saliwanchik
; STREET: 2421 N.W. 41st Street, Suite A-1
; CITY: Gainesville
; STATE: FL
; COUNTRY: US
; ZIP: 32606-6669
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/073,898
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/029,848
; FILING DATE: 30-OCT-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/960,780
; FILING DATE: 30-OCT-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Sanders, Jay M.
; REGISTRATION NUMBER: 39,355
; REFERENCE/DOCKET NUMBER: MA-708C1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 352-375-8100
; TELEFAX: 352-372-5800
; INFORMATION FOR SEQ ID NO: 38:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 357 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
```

```

;
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; ORIGINAL SOURCE:
; INDIVIDUAL ISOLATE: 196F3
US-09-073-898-38

Query Match      23.3%; Score 57; DB 3; Length 357;
Best Local Similarity 40.6%; Pred. No. 3.5;
Matches 13; Conservative 6; Mismatches 13; Indels 0; Gaps 0;

QY 17 LAESREQAKRLIMAGKVTLTNNSTTIPLRLEK 48
Db 63 LSTSSDQVFTLXGKIILNQSSMTFIRLEK 94

RESULT 14
US-09-850-351A-38
; Sequence 38, Application US/09850351A
; Patent No. 6656908
; GENERAL INFORMATION:
; APPLICANT: Feitelson, Jerald S.
; SCHNEPF, H. Ernest
; Narva, Kenneth E.
; Stockhoff, Brian A.
; Schmelts, James
; Loewer, David
; Dullum, Charles Joseph
; Muller-Cohn, Judy
; Stamp, Lisa
; Morrill, George
; TITLE OF INVENTION: No. 6656908el Pesticidal Toxins and Nucleotide
; Sequences Which Encode These Toxins
; NUMBER OF SEQUENCES: 144
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Saliwanchik, Lloyd & Saliwanchik
; STREET: 2421 N.W. 41st Street, Suite A-1
; CITY: Gainesville
; STATE: FL
; COUNTRY: US
; ZIP: 32606-6669
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/850,351A
; FILING DATE: 07-May-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 09/073,898
; FILING DATE: 06-MAY-1998
; APPLICATION NUMBER: US 08/960,780
; FILING DATE: 30-OCT-1997
; APPLICATION NUMBER: US 60/029,848
; FILING DATE: 30-OCT-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Sanders, Jay M.
; REGISTRATION NUMBER: 39,355
; REFERENCE/DOCKET NUMBER: WA-708CD1
; TELEPHONE: 352-375-8100
; TELEFAX: 352-372-5800
; INFORMATION FOR SEQ ID NO: 38:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 357 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; ORIGINAL SOURCE:
; INDIVIDUAL ISOLATE: 196F3
; SEQUENCE DESCRIPTION: SEQ ID NO: 38:

```

```

US-09-850-351A-38

Query Match      23.3%; Score 57; DB 4; Length 357;
Best Local Similarity 40.6%; Pred. No. 3.5;
Matches 13; Conservative 6; Mismatches 13; Indels 0; Gaps 0;

QY 17 LAESREQAKRLIMAGKVTLTNNSTTIPLRLEK 48
Db 63 LSTSSDQVFTLXGKIILNQSSMTFIRLEK 94

RESULT 15
US-09-134-000C-3991
; Sequence 3991, Application US/09134000C
; Patent No. 6617156
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; FILE OF INVENTION: ENTEROCOCCUS FAECALIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 032796-032
; CURRENT APPLICATION NUMBER: US/09/134,000C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/055,778
; PRIOR FILING DATE: 1997-08-15
; NUMBER OF SEQ ID NOS: 6812
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 3991
; LENGTH: 220
; TYPE: PRT
; ORGANISM: Enterococcus faecalis
US-09-134-000C-3991

Query Match      23.1%; Score 56.5; DB 4; Length 220;
Best Local Similarity 29.3%; Pred. No. 2.2;
Matches 12; Conservative 11; Mismatches 11; Indels 7; Gaps 1;

QY 6 VRADELVFLQGLAESREQAKRLIMAGKVTLTNNSTTIPRL 46
Db 160 VKDTSLLYVLGLD-----LLRAGKIAMSRDVTILVPLAL 193

RESULT 16
US-09-489-039A-8358
; Sequence 8358, Application US/09489039A
; Patent No. 6610836
; GENERAL INFORMATION:
; APPLICANT: Gary Breton et. al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
; FILE OF INVENTION: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 2709.2004001
; CURRENT APPLICATION NUMBER: US/09/489,039A
; CURRENT FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: US 60/117,747
; PRIOR FILING DATE: 1999-01-29
; NUMBER OF SEQ ID NOS: 14342
; SEQ ID NO 8358
; LENGTH: 476
; TYPE: PRT
; ORGANISM: Klebsiella pneumoniae
US-09-489-039A-8358

Query Match      23.1%; Score 56.5; DB 4; Length 476;
Best Local Similarity 31.5%; Pred. No. 6.1;
Matches 17; Conservative 6; Mismatches 22; Indels 9; Gaps 2;

QY 1 MAKHKVRADELVFLQGLAESREQAKRLIMAGKVTLTNNSTTIPL----RLEKPG 50
Db 398 MQMHHRRAEHVILAGTGQVTVNGKQFLIT-----ENQSTFIDIGAHSLENG 446

RESULT 17
US-09-252-991A-24632
; Sequence 24632, Application US/09252991A

```

```
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 24632
; LENGTH: 411
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-24632

Query Match      22.4%; Score 55; DB 4; Length 411;
Best Local Similarity 36.7%; Pred. No. 8.3;
Matches 18; Conservative 7; Mismatches 16; Indels 8; Gaps 2;

QY      7 RAEVLVFLQGLAESR-----EQAKRLIMAGKVTLTNNSTTPIRLKPG 50
      |||||:|||||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
Db      54 RADARMAMQGLARSQAHALDRARRRLRA---RLRLSGTAPARATGPG 99

RESULT 18
US-09-540-236-3643
; Patent No. 6673910
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO MORAXELLA CATAR
; TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 2709.2005-001
; CURRENT APPLICATION NUMBER: US/09/540,236
; CURRENT FILING DATE: 2000-04-04
; NUMBER OF SEQ ID NOS: 3840
; SEQ ID NO 3643
; LENGTH: 153
; TYPE: PRT
; ORGANISM: M.catarrhalis
US-09-540-236-3643

Query Match      22.0%; Score 54; DB 4; Length 153;
Best Local Similarity 32.8%; Pred. No. 3.2;
Matches 14; Conservative 7; Mismatches 22; Indels 0; Gaps 0;

QY      5 KVRADLVFLQGLAESREQAKRLIMAGKVTLTNNSTTPIRLLE 47
      |||||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
Db      28 KVRADKWLAAARFRTRSLAKEALEAGKVHMGAKIKVSKELQ 70

RESULT 19
US-09-723-368-2
; Sequence 2, Application US/09723368
; Patent No. 6641818
; GENERAL INFORMATION:
; APPLICANT: NORTHWESTERN UNIVERSITY
; APPLICANT: SPEAR, Patricia G.
; APPLICANT: WARNER, Morgyn S.
; APPLICANT: GERAGHTY, Robert G.
; APPLICANT: MARTINEZ, Wanda M.
; APPLICANT: MONTGOMERY, Rebecca I.
; APPLICANT: COHEN, Gary H.
; APPLICANT: EISENBERG, Roselyn J.
; APPLICANT: WHITBECK, Charles J.
; APPLICANT: KRUMMENACHER, Claude
; APPLICANT: UNIVERSITY OF PENNSYLVANIA
; TITLE OF INVENTION: CELLULAR PROTEINS WHICH MEDIATE HERPESVIRUS ENTRY
; FILE REFERENCE: 200290.0050/201
```

```
; CURRENT APPLICATION NUMBER: US/09/723,368
; CURRENT FILING DATE: 2000-11-28
; PRIOR APPLICATION NUMBER: U.S. 60/087,862
; PRIOR FILING DATE: 1998-06-03
; PRIOR APPLICATION NUMBER: PCT/US99/12235
; PRIOR FILING DATE: 1999-06-02
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 479
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-723-368-2
```

```
Query Match      22.0%; Score 54; DB 4; Length 479;
Best Local Similarity 33.3%; Pred. No. 14;
Matches 16; Conservative 11; Mismatches 15; Indels 6; Gaps 2;
```

```
QY      3 KHKVRADELVFLQGLAESREQAKRLIMAGKVTLTNNSTTPIRLKPG 50
      |||||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
Db      147 KGSVRG--MTWLRVIAPKQAE-----AQKVTFSQDPTTVALCISKEG 188
```

```
RESULT 20
US-09-252-991A-19202
; Sequence 19202, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 19202
; LENGTH: 480
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-19202
```

```
Query Match      21.8%; Score 53.5; DB 4; Length 480;
Best Local Similarity 32.7%; Pred. No. 17;
Matches 17; Conservative 8; Mismatches 16; Indels 11; Gaps 3;
```

```
QY      4 HKVRADELVFLQGLAESREQAKRLIMAGKVTLTNNSTTPIPL-----RLKPG 50
      |||||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
Db      404 HHRSEHWIVVQGMARVTNGDGLRVN-----SNSTYIPAGHRHLENPG 449
```

```
RESULT 21
US-09-732-210-77
; Sequence 77, Application US/09732210
; Patent No. 6573361
; GENERAL INFORMATION:
; APPLICANT: Bunkers, Greg J.
; APPLICANT: Liang, Jihong
; APPLICANT: Mittanck, Cindy A.
; APPLICANT: Seale, Jeffrey W.
; APPLICANT: Wu, Yonnie S.
; TITLE OF INVENTION: Anti-fungal Proteins and Methods for Their Use
; FILE REFERENCE: 38-21(15036)B
; CURRENT APPLICATION NUMBER: US/09/732,210
; CURRENT FILING DATE: 2000-12-07
; PRIOR APPLICATION NUMBER: US 60/169,513
; PRIOR FILING DATE: 1999-12-07
; PRIOR APPLICATION NUMBER: US 60/169,340
; PRIOR FILING DATE: 1999-12-07
; NUMBER OF SEQ ID NOS: 1753
```

```

; SEQ ID NO 77
; LENGTH: 17
; TYPE: PRT
; ORGANISM:
US-09-732-210-

```

Query Match 21.6%; Score 53; DB 4; Length 171;
Best Local Similarity 27.3%; Pred.No.5.2;
Matches 12; Conservative 11; Mismatches 21; Indels

QY 7 RADELVFLOGLAERSQAERLINMAGKVTITNNSTTPTLELEKPG 50
:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 104 RLOTLVHKSLARTVEAROMVTHGIALDGRKDIDAPGVIVRKG 147

```

RESULT 22
US-09-328-352-4325
; Sequence 4325, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
; APPLICANT: GARY L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND
; TITLE OF INVENTION: BAUMANNII FOR DNA
; FILE REFERENCE: GTC99-03PA
; CURRENT APPLICATION NUMBER: US/09/328
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 4325
; LENGTH: 212
; TYPE: prt
; ORGANISM: Acinetobacter baumannii
US-09-328-352-4325

```

Query Match 21.6%; Score 53; DB 4; Length 212;
Best Local Similarity 27.8%; Pred.No. 6.9;
Matches 10; Conservative 8; Mismatches 18; Indels

QY 7 RADELVFLGGLAFESREQKELIMAGKVTLNNSTTI 42
 :::|:::|:
Db 103 RLNDVVYRNGFGFSTRAEAROLVSHRSITLNGRRVNI 138

```

RESULT 23
US-09-199-637A-209
: Sequence 209, Application US/09199637A
: Patent No. 6355411
: GENERAL INFORMATION:
: APPLICANT: Ausubel, Frederick
: APPLICANT: Goodman, Howard M.
: APPLICANT: Rahme, Laurence G.
: APPLICANT: Mahajan-Miklos, Shalina
: APPLICANT: Tan, Man-Wah
: APPLICANT: Cao, Hui
: APPLICANT: Drenkard, Eliana
: APPLICANT: Tsengalis, John
: TITLE OF INVENTION: VIRULENCE-ASSOCI-
: TITLE OF INVENTION: SEQUENCES AND US-
: FILE REFERENCE: 00786/361002
: CURRENT APPLICATION NUMBER: US/09/199-
: CURRENT FILING DATE: 1998-11-25
: PRIOR APPLICATION NUMBER: 60/086,517
: PRIOR FILING DATE: 1997-11-25
: NUMBER OF SEQ ID NOS: 437
: SOFTWARE: FastSeq for Windows Versio
: SEQ ID NO 209
: LENGTH: 252
: TYPE: PRT
: ORGANISM: Pseudomonas aeruginosa
US-09-199-637A-209

```

Query Match 21.6%; Score 53; DB 4; Length 252;
Best Local Similarity 52.4%; Pred. No. 8.6;
Matches 11; Conservative 5; Mismatches 5; Indels

QY 8 ADELVFLQLAESREQAKRLI 28
 |||:::| | | | | :
Dh 158 ADOLLVVQOAAHVROAKRLL 178

RESULT 24

US-09-252-991A-29761

Sequence 29761, Application US/09252991A

Patent No. 6551795

GENERAL INFORMATION:

APPLICANT: Marc J. Rubenfield et al.

TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS

TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS

FILE REFERENCE: 107196.136

CURRENT APPLICATION NUMBER: US/09/252,991A

CURRENT FILING DATE: 1999-02-18

PRIOR APPLICATION NUMBER: US 60/074,788

PRIOR FILING DATE: 1998-02-18

PRIOR APPLICATION NUMBER: US 60/094,190

PRIOR FILING DATE: 1998-07-27

NUMBER OF SEQ ID NOS: 33142

SEQ ID NO 29761

LENGTH: 330

TYPE: PRT

ORGANISM: Pseudomonas aeruginosa

US-09-252-991A-29761

Query Match	21.6%	Score 53;	DB 4;	Length 330;
Best Local Similarity	52.4%	Pred. No. 12;		
Matches	11.	Conservative	5;	Mismatches
			5;	Indels
			0;	Gaps

QY 8 ADELVFLOGLAESREQAKRLI 28
|||: : | | : : |||:
Db 236 ADOLLVVOOAAHVDRDOAKRLI 256

RESULT 25
US-09-199-637A-107
; Sequence 107, Application US/09199637A
; Patent No. 6355411
; GENERAL INFORMATION:
; APPLICANT: Ausubel, Frederick
; APPLICANT: Goodman, Howard M.
; APPLICANT: Rahme, Laurence G.
; APPLICANT: Mahajan-Miklos, Shalina
; APPLICANT: Tan, Man-Wah
; APPLICANT: Cao, Hui
; APPLICANT: Drenkard, Eliana
; APPLICANT: Tsongalis, John
; TITLE OF INVENTION: VIRULENCE-ASSOCIATED NUCLEIC ACID
; TITLE OF INVENTION: SEQUENCES AND USES THEREOF
; FILE REFERENCE: 00786/361002
; CURRENT APPLICATION NUMBER: US/09/199,637A
; CURRENT FILING DATE: 1998-11-25
; PRIOR APPLICATION NUMBER: 60/066,517
; PRIOR FILING DATE: 1997-11-25
; NUMBER OF SEQ ID NOS: 437
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 107
; LENGTH: 427
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-199-637A-107

Query Match 21.6%; Score 53; DB 4; Length 427;
Best Local Similarity 52.4%; Pred. No. 17;
Matches 11; Conservative 5; Mismatches 5; Indels

QY 8 ADELVFLOGLAESREQAKRLI 28
 ||| : : | - : : ||| :
Db 333 ADOLLVVOAAHVDDQAKRLL 353

[illegible]

; OPERATING SYSTEM: <Unknown>
; SOFTWARE: ASCII
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/107,532A
; FILING DATE: 30-Jun-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/085,598
; FILING DATE: 14 May 1998
; APPLICATION NUMBER: 60/051571
; FILING DATE: July 2, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Ariniello, Pamela Deneke
; REGISTRATION NUMBER: 40,489
; REFERENCE/DOCKET NUMBER: GTC-012
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (781)893-5007
; TELEFAX: (781)893-8277
; INFORMATION FOR SEQ ID NO: 6634:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 240 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: YES
; ORIGINAL SOURCE:
; ORGANISM: Enterococcus faecium
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (B) LOCATION 1...240
; SEQUENCE DESCRIPTION: SEQ ID NO: 6634:
US-09-107-532A-6634

Query Match 20.4%; Score 50; DB 4; Length 240;
Best Local Similarity 31.8%; Pred. No. 23;
Matches 14; Conservative 6; Mismatches 24; Indels 0; Gaps 0;
QY 5 KVRADLVFLQGLAESREQAKRLIMAGKVTLTNNSTTIPLRLPK 48
Db 165 EVNTDDLITMLRLDQFMQNGMYIKADKIDLVNDIRFSLYLPK 208

RESULT 39
US-09-134-000C-5324
; Sequence 5324, Application US/09134000C
; Patent No. 6617156
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; FILE REFERENCE: 032796-032
; CURRENT APPLICATION NUMBER: US/09/134,000C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/055,778
; PRIOR FILING DATE: 1997-08-15
; NUMBER OF SEQ ID NOS: 6812
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 5324
; LENGTH: 315
; TYPE: PRT
; ORGANISM: Enterococcus faecalis
US-09-134-000C-5324

Query Match 20.4%; Score 50; DB 4; Length 315;
Best Local Similarity 40.5%; Pred. No. 33;
Matches 17; Conservative 3; Mismatches 22; Indels 0; Gaps 0;

QY 5 KVRADLVFLQGLAESREQAKRLIMAGKVTLTNNSTTIPLRL 46
Db 4 KKEAIFMIELIATAESVEQAVELLATGVDLTLYIGETFFGLRL 45

RESULT 40
US-09-107-532A-4924

; Sequence 4924, Application US/09107532A
; Patent No. 6583275
; GENERAL INFORMATION:
; APPLICANT: Lynn A Doucette-Stamm and David Bush
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: GENOME THERAPEUTICS CORPORATION
; STREET: 100 Beaver Street
; CITY: Waltham
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02354
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD-ROM ISO9660
; OPERATING SYSTEM: <Unknown>
; SOFTWARE: ASCII
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/107,532A
; FILING DATE: 30-Jun-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/085,598
; FILING DATE: 14 May 1998
; APPLICATION NUMBER: 60/051571
; FILING DATE: July 2, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Ariniello, Pamela Deneke
; REGISTRATION NUMBER: 40,489
; REFERENCE/DOCKET NUMBER: GTC-012
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (781)893-5007
; TELEFAX: (781)893-8277
; INFORMATION FOR SEQ ID NO: 4924:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 316 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: YES
; ORIGINAL SOURCE:
; ORGANISM: Enterococcus faecium
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (B) LOCATION 1...316
; SEQUENCE DESCRIPTION: SEQ ID NO: 4924:
US-09-107-532A-4924
Query Match 20.4%; Score 50; DB 4; Length 316;
Best Local Similarity 31.2%; Pred. No. 33;
Matches 15; Conservative 10; Mismatches 21; Indels 2; Gaps 2;
QY 4 HKVRADLVFLQGLAESREQAKRLIMAGK-VTLTNNSTTIPLRLPK 50
Db 169 HKIVGDAARNGASLAELKELGEKVQVQATKTIGLALRAATVP-EVGKPG 215
Search completed: July 7, 2004, 16:58:11
Job time : 6.48173 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 7, 2004, 16:56:51 ; Search time 14.1196 Seconds
(without alignments)
1102.311 Million cell updates/sec

Title: US-10-009-919A-1_COPY_1_50

Perfect score: 245

Sequence: 1 MAKHVRADLVFLQGLAES.....GKVTLTNNSTPIRLKPG 50

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1276540 segs, 311283816 residues

Total number of hits satisfying chosen parameters: 1276540

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:*

- 1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
- 2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep.*
- 3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep.*
- 4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep.*
- 5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*
- 6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep.*
- 7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*
- 8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
- 9: /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep.*
- 10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep.*
- 11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep.*
- 12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep.*
- 13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep.*
- 14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep.*
- 15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep.*
- 16: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep.*
- 17: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep.*
- 18: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	106.5	43.5	276	15	US-10-264-213-162
2	95	38.8	270	9	US-09-765-272-64
3	87	35.5	191	12	US-10-282-122A-53447
4	75	30.6	57	11	US-09-930-512-33
5	75	30.6	209	13	US-10-013-379-27
6	74	30.2	137	9	US-09-939-980-431
7	74	30.2	195	14	US-10-138-701-46
8	74	30.2	200	9	US-09-815-242-5404
9	74	30.2	200	9	US-09-815-242-12578
10	74	30.2	200	9	US-09-815-242-12778
11	74	30.2	200	9	US-09-908-931B-11
12	74	30.2	200	12	US-10-282-122A-44034
13	74	30.2	200	16	US-10-332-964-12
14	73	29.8	200	12	US-10-282-122A-45742
15	73	29.8	200	12	US-10-282-122A-71012

16	29.4	42	11	US-09-930-512-34	Sequence 34, Appl
17	29.0	201	9	US-09-815-242-5030	Sequence 5030, Ap
18	29.0	203	9	US-09-815-242-10879	Sequence 10879, A
19	29.0	203	12	US-10-282-122A-42445	Sequence 42445, A
20	28.6	201	9	US-09-738-626-4126	Sequence 4126, Ap
21	28.6	201	12	US-10-282-122A-53648	Sequence 53648, A
22	28.2	1147	15	US-10-369-493-22316	Sequence 22316, A
23	28.2	1147	16	US-10-618-581-2	Sequence 2, Appli
24	27.8	205	12	US-10-282-122A-64188	Sequence 64188, A
25	27.3	200	12	US-10-282-122A-71298	Sequence 71298, A
26	27.3	201	12	US-10-282-122A-48598	Sequence 48598, A
27	27.3	205	12	US-10-282-122A-63563	Sequence 63563, A
28	26.9	200	12	US-10-282-122A-60489	Sequence 60489, A
29	26.9	202	12	US-10-282-122A-68872	Sequence 76872, A
30	26.1	203	9	US-09-815-242-13240	Sequence 13240, A
31	26.1	203	12	US-10-282-122A-73698	Sequence 73698, A
32	26.1	203	12	US-10-282-122A-74814	Sequence 74814, A
33	26.1	203	16	US-10-474-776-751	Sequence 751, App
34	25.7	209	12	US-10-282-122A-47209	Sequence 47209, A
35	25.7	235	12	US-10-335-977-7740	Sequence 7740, Ap
36	25.7	235	12	US-10-335-977-7741	Sequence 7741, Ap
37	25.7	246	12	US-10-335-977-7742	Sequence 7742, Ap
38	24.9	201	12	US-10-282-122A-61700	Sequence 61700, A
39	24.5	203	12	US-10-282-122A-72484	Sequence 72484, A
40	24.3	582	12	US-10-282-122A-74807	Sequence 74807, A
41	24.3	1510	12	US-10-282-122A-49816	Sequence 49816, A
42	24.1	206	12	US-10-282-122A-68056	Sequence 68056, A
43	24.1	208	9	US-09-815-242-11425	Sequence 11425, A
44	24.1	208	12	US-10-282-122A-58944	Sequence 58944, A
45	24.1	208	12	US-10-335-977-7159	Sequence 7159, Ap

ALIGNMENTS

RESULT 1

US-10-264-213-162
; Sequence 162, Application US/10264213
; Publication No. US20040009490A1
; GENERAL INFORMATION:
; APPLICANT: Glenn, Matthew
; APPLICANT: Havukkala, Ilkka J
; APPLICANT: Lubbers, Mark William
; APPLICANT: Dekker, James
; TITLE OF INVENTION: Polynucleotides, materials incorporating
; TITLE OF INVENTION: them, and methods for using them.
; FILE REFERENCE: 11000.1843c3
; CURRENT APPLICATION NUMBER: US/10/264,213
; CURRENT FILING DATE: 2002-10-03
; NUMBER OF SEQ ID NOS: 253
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 162
; LENGTH: 276
; TYPE: PRT
; ORGANISM: Lactobacillus rhamnosus
US-10-264-213-162

Query Match 43.5%; Score 106.5; DB 15; Length 276;
Best Local Similarity 56.5%; Pred. No. 1.7e-06;
Matches 26; Conservative 4; Mismatches 11; Indels 5; Gaps 1;

QY 5 KVRADLVFLQGLAESREQAKRLIMAGKVTLTNNSTPIRLKPG 50

Db 6 KERVVDLVFLQGLFESFEQAKRSMAGEVVDQNNQ-----RLDKPG 46

RESULT 2

US-09-765-272-64
; Sequence 64, Application US/09765272
; Patent No. US20020061545A1
; GENERAL INFORMATION:
; APPLICANT: Choi et. al.
; TITLE OF INVENTION: Streptococcus pneumoniae Antigen and Vaccines

```

;
; NUMBER OF SEQUENCES: 452
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
;
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/765,272
; FILING DATE: 22-Jan-2001
; CLASSIFICATION: <Unknown>
;
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/961,083
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Brookes, A. Anders
; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: PB340P2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
;
; INFORMATION FOR SEQ ID NO: 64:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 270 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
;
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 64:
US-09-765-272-64

Query Match 38.8%; Score 95; DB 9; Length 270;
Best Local Similarity 50.0%; Pred. No. 7.1e-05;
Matches 23; Conservative 4; Mismatches 15; Indels 4; Gaps 1;

QY 5 KVRDELVFLQGLAESREQAKRLIMAGKVTILTNNSSTIPLRLKPG 50
Db 2 KERVDVLAYKQGLFETREQAKRGVMAGLVAVLNGE----RFDKPG 43

RESULT 3
US-10-282-122A-53447
; Sequence 53447, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26

```

```

;
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 53447
; LENGTH: 191
; TYPE: PRT
; ORGANISM: Clostridium difficile
; US-10-282-122A-53447

Query Match 35.5%; Score 87; DB 12; Length 191;
Best Local Similarity 40.9%; Pred. No. 0.00062;
Matches 18; Conservative 8; Mismatches 18; Indels 0; Gaps 0;

QY 7 RADELVFLQGLAESREQAKRLIMAGKVTILTNNSSTIPLRLKPG 50
Db 83 RLDNVVYRNGGLASSRKEARQLVTHGHFTLNGKNKVDIPSLIVKVG 126

RESULT 4
US-09-930-512-33
; Sequence 33, Application US/09930512
; Publication No. US20040010118A1
; GENERAL INFORMATION:
; APPLICANT: Zerhusen, Bryan D
; APPLICANT: Padigaru, Muralidhara
; APPLICANT: Spytek, Kimberly
; APPLICANT: Spaderna, Steven
; APPLICANT: Gangolli, Esha A
; APPLICANT: Rastelli, Luca
; APPLICANT: Burgess, Catherine E
; APPLICANT: Majumder, Kumud
; APPLICANT: Shimkets, Richard
; APPLICANT: Mishra, Vishnu
; APPLICANT: Vernet, Corine
; APPLICANT: Szekeres, Edward S
; APPLICANT: Grosse, William M
; APPLICANT: Alsobrook II, John P
; APPLICANT: Liu, Xiaohong
; APPLICANT: Gerlach, Valerie L
; APPLICANT: Ellerman, Karen
; APPLICANT: Smithson, Glennda
; APPLICANT: Feyman, John
; APPLICANT: Stone, David
; APPLICANT: MacDougall, John
; TITLE OF INVENTION: No. US20040010118A1e1 Proteins and Nucleic Acids Encoding Same
; FILE REFERENCE: 21402-091
; CURRENT APPLICATION NUMBER: US/09/930,512
; CURRENT FILING DATE: 2001-08-15
; PRIOR APPLICATION NUMBER: 60/225,692
; PRIOR FILING DATE: 2000-08-16
; PRIOR APPLICATION NUMBER: 60/225,837
; PRIOR FILING DATE: 2000-08-16
; PRIOR APPLICATION NUMBER: 60/225,693
; PRIOR FILING DATE: 2000-08-16
; PRIOR APPLICATION NUMBER: 60/226,236
; PRIOR FILING DATE: 2000-08-18
; PRIOR APPLICATION NUMBER: 60/226,353
; PRIOR FILING DATE: 2000-08-18
; PRIOR APPLICATION NUMBER: 60/227,085

```

```
; PRIOR FILING DATE: 2000-08-22
; PRIOR APPLICATION NUMBER: 60/227,395
; PRIOR FILING DATE: 2000-08-23
; PRIOR APPLICATION NUMBER: 60/227,492
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: 60/227,600
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: 60/275,952
; PRIOR FILING DATE: 2001-03-14
; NUMBER OF SEQ ID NOS: 115
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 33
; LENGTH: 57
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-930-512-33

Query Match      30.6%; Score 75; DB 11; Length 57;
Best Local Similarity 45.7%; Pred. No. 0.0065;
Matches 16; Conservative 5; Mismatches 14; Indels 0; Gaps 0;

Qy      16 GLAESRQAKRLIMAGKVTLTNNSTTIPRLKPG 50
Db      6 GLAPRSQARKLTBQGRVKVNGKXVTDPYSYIVKPG 40

RESULT 5
US-10-013-379-27
; Sequence 27, Application US/10013379
; Publication No. US20020188108A1
; GENERAL INFORMATION:
; APPLICANT: The Regents of the University of California
; APPLICANT: Cate, Jamie H.
; APPLICANT: No. US20020188108A1, Harry F.
; APPLICANT: Yusupov, Marat M.
; APPLICANT: Yusupova, Guinara ZH
; APPLICANT: Baucom, Albion
; APPLICANT: Lancaster, Laura
; APPLICANT: Dallas, Anne
; TITLE OF INVENTION: X-RAY CRYSTAL STRUCTURES OF FUNCTIONAL
; TITLE OF INVENTION: COMPLEXES OF THE BACTERIAL RIBOSOME CONTAINING TRANSFER RNA
; TITLE OF INVENTION: AND MODEL MESSENGER RNAS
; FILE REFERENCE: 19629-7010
; CURRENT APPLICATION NUMBER: US/10/013,379
; CURRENT FILING DATE: 2002-03-18
; PRIOR APPLICATION NUMBER: US 60/254,603
; PRIOR FILING DATE: 2000-12-09
; PRIOR APPLICATION NUMBER: US 60/278,013
; PRIOR FILING DATE: 2001-03-22
; PRIOR APPLICATION NUMBER: US 60/294,394
; PRIOR FILING DATE: 2001-05-30
; NUMBER OF SEQ ID NOS: 54
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 27
; LENGTH: 209
; TYPE: PRT
; ORGANISM: Thermus thermophilus
; FEATURE:
; OTHER INFORMATION: 30S ribosomal protein S4
; OTHER INFORMATION: IgiXG
US-10-013-379-27

Query Match      30.6%; Score 75; DB 13; Length 209;
Best Local Similarity 34.1%; Pred. No. 0.035;
Matches 15; Conservative 9; Mismatches 20; Indels 0; Gaps 0;

Qy      7 RADELVFQGLAESRQAKRLIMAGKVTLTNNSTTIPRLKPG 50
Db     100 RLDNVVYELGFAVSRQARQLVRHGHIIVNGRRVDLPYSYVRPG 143

RESULT 6
US-09-939-980-431
; Sequence 431, Application US/09939980
; Patent No. US20020082234A1
; GENERAL INFORMATION:
; APPLICANT: Black, Michael
; Burnham, Martin
; Hodgson, John
; Knowles, David
; Lonetto, Michael
; Nicholas, Richard
; Pratt, Julie
; Reichard, Richard
; Rosenberg, Martin
; Ward, Judith
; TITLE OF INVENTION: NO. US20020082234A1, Polypeptides and Their Uses
; Polypeptides and Their Uses
; NUMBER OF SEQUENCES: 534
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SmithKline Beecham Corporation
; STREET: 709 Swedeland Road
; CITY: King of Prussia
; STATE: PA
; COUNTRY: USA
; ZIP: 19406-0939
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/939,980
; FILING DATE: 27-Aug-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/936,165
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Gimmi, Edward R
; REGISTRATION NUMBER: 38,891
; REFERENCE/DOCKET NUMBER: P50549
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 610-270-4478
; TELEFAX: 610-270-5090
; TELEX: <Unknown>
; INFORMATION FOR SEQ ID NO: 431:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 137 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: Protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 431:
US-09-939-980-431

Query Match      30.2%; Score 74; DB 9; Length 137;
Best Local Similarity 36.4%; Pred. No. 0.028;
Matches 16; Conservative 8; Mismatches 20; Indels 0; Gaps 0;

Qy      7 RADELVFQGLAESRQAKRLIMAGKVTLTNNSTTIPRLKPG 50
Db     14 RLDVAVYSLGLARTRQARQLVNHGHILVDGKRVDPISYSVKPG 57

RESULT 7
US-10-138-701-46
; Sequence 46, Application US/10138701
; Publication No. US20030186364A1
; GENERAL INFORMATION:
; APPLICANT: Human Genome Sciences, Inc. et al.
; TITLE OF INVENTION: Staphylococcus aureus genes and polypeptides
; FILE REFERENCE: PB484
; CURRENT APPLICATION NUMBER: US/10/138,701
; CURRENT FILING DATE: 2002-05-06
; PRIOR APPLICATION NUMBER: US/09/512,255A
```

;
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/098,964
; PRIOR FILING DATE: 1998-09-01
; PRIOR APPLICATION NUMBER: US 60/009,861
; PRIOR FILING DATE: 1996-01-05
; PRIOR APPLICATION NUMBER: PCT/ US99/19726
; PRIOR FILING DATE: 1999-08-31
; PRIOR APPLICATION NUMBER: US 08/956,171
; PRIOR FILING DATE: 1997-10-20
; NUMBER OF SEQ ID NOS: 61
; SOFTWARE: Patentin version 3.0
; SEQ ID NO 46
; LENGTH: 195
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
US-10-138-701-46

Query Match 30.2%; Score 74; DB 14; Length 195;
Best Local Similarity 36.4%; Pred. No. 0.046;
Matches 16; Conservative 8; Mismatches 20; Indels 0; Gaps 0;
QY 7 RADELVFOGLAESRQAKRLIMAGKVTLTNNSTTIPRLKPG 50
|||:|||||:|||||:|:|
Db 93 RLDVVYSLGLARTRQARQLVNHGHILVDGKRDIPSVKPG 136

RESULT 8
US-09-815-242-5404
; Sequence 5404, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 5404
; LENGTH: 200
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
US-09-815-242-5404

Query Match 30.2%; Score 74; DB 9; Length 200;
Best Local Similarity 36.4%; Pred. No. 0.046;
Matches 16; Conservative 8; Mismatches 20; Indels 0; Gaps 0;
QY 7 RADELVFOGLAESRQAKRLIMAGKVTLTNNSTTIPRLKPG 50
|||:|||||:|||||:|:|
Db 93 RLDVVYSLGLARTRQARQLVNHGHILVDGKRDIPSVKPG 136

RESULT 9

US-09-815-242-12578
; Sequence 12578, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 12578
; LENGTH: 200
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
US-09-815-242-12578

Query Match 30.2%; Score 74; DB 9; Length 200;
Best Local Similarity 36.4%; Pred. No. 0.046;
Matches 16; Conservative 8; Mismatches 20; Indels 0; Gaps 0;

QY 7 RADELVFOGLAESRQAKRLIMAGKVTLTNNSTTIPRLKPG 50
|||:|||||:|||||:|:|
Db 93 RLDVVYSLGLARTRQARQLVNHGHILVDGKRDIPSVKPG 136

RESULT 10

US-09-815-242-12778
; Sequence 12778, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23

```
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12778
; LENGTH: 200
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
US-09-815-242-12778

Query Match      30.2%; Score 74; DB 9; Length 200;
Best Local Similarity 36.4%; Pred. No. 0.046;
Matches 16; Conservative 8; Mismatches 20; Indels 0; Gaps 0;

Qy      7 RADELVFLQGLAESREQAQRILIMAGKVTLTNNSTTIPLRLEKPG 50
      |||::|||::|||::|||::|||::|||::|||::|||::|||
Db      93 RLDVVYSLGLARTRRQARQLVNHGHILVDGKRDIPSYSVKPG 136

RESULT 11
US-09-908-931B-11
; Sequence 11, Application US/09908931B
; Patent No. US20020164756A1
; GENERAL INFORMATION:
; APPLICANT: Pearson, James D.
; APPLICANT: Slightom, Jerry
; APPLICANT: Chosay, John G.
; APPLICANT: Shinabarger, Dean L.
; TITLE OF INVENTION: Complete Nucleotide Sequence of Staphylococcus aureus
; TITLE OF INVENTION: Ribosomal Protein Gene, S16 and Methods for the
; TITLE OF INVENTION: Identification of Antibacterial Substances
; FILE REFERENCE: S16 ribosomal protein
; CURRENT APPLICATION NUMBER: US/09/908,931B
; CURRENT FILING DATE: 2001-07-19
; NUMBER OF SEQ ID NOS: 58
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 11
; LENGTH: 200
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
US-09-908-931B-11

Query Match      30.2%; Score 74; DB 9; Length 200;
Best Local Similarity 36.4%; Pred. No. 0.046;
Matches 16; Conservative 8; Mismatches 20; Indels 0; Gaps 0;

Qy      7 RADELVFLQGLAESREQAQRILIMAGKVTLTNNSTTIPLRLEKPG 50
      |||::|||::|||::|||::|||::|||::|||::|||
Db      93 RLDVVYSLGLARTRRQARQLVNHGHILVDGKRDIPSYSVKPG 136

RESULT 12
US-10-282-122A-44034
; Sequence 44034, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
```

```
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 44034
; LENGTH: 200
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
US-10-282-122A-44034

Query Match      30.2%; Score 74; DB 12; Length 200;
Best Local Similarity 36.4%; Pred. No. 0.046;
Matches 16; Conservative 8; Mismatches 20; Indels 0; Gaps 0;

Qy      7 RADELVFLQGLAESREQAQRILIMAGKVTLTNNSTTIPLRLEKPG 50
      |||::|||::|||::|||::|||::|||::|||::|||
Db      93 RLDVVYSLGLARTRRQARQLVNHGHILVDGKRDIPSYSVKPG 136

RESULT 13
US-10-332-964-12
; Sequence 12, Application US/10332964
; Publication No. US20040097718A1
; GENERAL INFORMATION:
; APPLICANT: Pearson, James D
; APPLICANT: Slightom, Jerry
; APPLICANT: Chosay, John G.
; APPLICANT: McCroskey, Mark C.
; APPLICANT: Shinabarger, Dean L.
; APPLICANT: Wilcox, Sheri
; TITLE OF INVENTION: Complete Nucleotide Sequence of Staphylococcus aureus
; TITLE OF INVENTION: Ribosomal Protein Gene, S20 and Methods for the
; TITLE OF INVENTION: Identification of Antibacterial Substances
; FILE REFERENCE: Complete Nuc Seq S20
; CURRENT APPLICATION NUMBER: US/10/332,964
; CURRENT FILING DATE: 2003-10-29
; NUMBER OF SEQ ID NOS: 62
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 12
; LENGTH: 200
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
US-10-332-964-12

Query Match      30.2%; Score 74; DB 16; Length 200;
Best Local Similarity 36.4%; Pred. No. 0.046;
Matches 16; Conservative 8; Mismatches 20; Indels 0; Gaps 0;
```

Qy 7 RADELVFLQGLAESREQAKELIWAGKVTLLNNSTTIPRLKEPG 50
| | : : | | : : | | : : | | : : | |
Db 93 RLDAVYSLGLARTTRQARQLNVNHHILVDGKRVDPISYVKPG 136

```

RESULT 14
US-10-282-122A-45742
; Sequence 45742, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A

```

```

: APPLICANT: Haselbeck, Robert
: APPLICANT: Ohlsen, Karl
: APPLICANT: Zyskind, Judith
: APPLICANT: Wall, Daniel
: APPLICANT: Trawick, John
: APPLICANT: Carr, Grant
: APPLICANT: Yamamoto, Robert
: APPLICANT: Forsyth, R.
: APPLICANT: Xu, H.
: TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
: FILE REFERENCE: ELITRA.034A
: CURRENT APPLICATION NUMBER: US/10/282,122A
: CURRENT FILING DATE: 2003-02-20
: PRIOR APPLICATION NUMBER: 60/191,078
: PRIOR FILING DATE: 2000-03-21
: PRIOR APPLICATION NUMBER: 60/206,848
: PRIOR FILING DATE: 2000-05-23
: PRIOR APPLICATION NUMBER: 60/207,727
: PRIOR FILING DATE: 2000-05-26
: PRIOR APPLICATION NUMBER: 60/230,335
: PRIOR FILING DATE: 2000-09-06
: PRIOR APPLICATION NUMBER: 60/230,347
: PRIOR FILING DATE: 2000-09-09
: PRIOR APPLICATION NUMBER: 60/242,578
: PRIOR FILING DATE: 2000-10-23
: PRIOR APPLICATION NUMBER: 60/253,625
: PRIOR FILING DATE: 2000-11-27
: PRIOR APPLICATION NUMBER: 60/257,931
: PRIOR FILING DATE: 2000-12-22
: PRIOR APPLICATION NUMBER: 60/267,636
: PRIOR FILING DATE: 2001-02-09
: PRIOR APPLICATION NUMBER: 60/269,308
: PRIOR FILING DATE: 2001-02-16
: Remaining Prior Application data removed - See File Wrapper or PALM.
: NUMBER OF SEQ ID NOS: 78614
: SOFTWARE: PatentIn version 3.1
: SEQ ID NO 71012
: LENGTH: 200
: TYPE: PRT
: ORGANISM: Staphylococcus epidermidis
: US-10-282-122A-71012

```

```
Query Match      29.8%; Score 73; DB 12; Length 200;
Best Local Similarity 36.4%; Pred.No. 0.063;
Matches 16; Conservative 8; Mismatches 20; Indels 0; Gaps 0;

Qy 7 RADELVLQGLAIREQAQLIMAGKVTLTNNSTTPIRLKEPG 50
          ||| | | | | | | | | | | | | | | |
Db 93 RLDAVYSLGLATRRAROLVNHGHIIEVDGGRDVDFSYLKPQ 136
```

```

; LENGTH: 200
; TYPE: PRF
; ORGANISM: Bacillus anthracis
US-10-282-122A-45742

Query Match          29.8%; Score 73; DB 12; Length 200;
Best Local Similarity 34.8%; Pred.No. 0.063;
Matches 16; Conservative 8; Mismatches 22; Indels 0; Gaps 0;

QY      5 KVRADIEIVFFLQGLAESFEQAKRITMAGKVTITNNSTTIPRLKXPG 50
       : |||: |||: |::|: |::|: |::|: |||
Db      91 EARLDNLVYRMGLARTRAARQLVNHGHIWVGARVDIPSRYRVFG 136

RESULT 15
US-10-282-122A-71012
; Sequence 71012, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
```



```
; APPLICANT: Smithson, Glennda
; APPLICANT: Peyman, John
; APPLICANT: Stone, David
; APPLICANT: MacDougall, John
; TITLE OF INVENTION: No. US20040010118a1el Proteins and Nucleic Acids Encoding Same
; FILE REFERENCE: 21402-091
; CURRENT APPLICATION NUMBER: US/09/930,512
; CURRENT FILING DATE: 2001-08-15
; PRIOR APPLICATION NUMBER: 60/225,692
; PRIOR FILING DATE: 2000-08-16
; PRIOR APPLICATION NUMBER: 60/225,837
; PRIOR FILING DATE: 2000-08-16
; PRIOR APPLICATION NUMBER: 60/225,693
; PRIOR FILING DATE: 2000-08-16
; PRIOR APPLICATION NUMBER: 60/226,236
; PRIOR FILING DATE: 2000-08-18
; PRIOR APPLICATION NUMBER: 60/226,353
; PRIOR FILING DATE: 2000-08-18
; PRIOR APPLICATION NUMBER: 60/227,085
; PRIOR FILING DATE: 2000-08-22
; PRIOR APPLICATION NUMBER: 60/227,395
; PRIOR FILING DATE: 2000-08-23
; PRIOR APPLICATION NUMBER: 60/227,492
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: 60/227,600
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: 60/275,952
; PRIOR FILING DATE: 2001-03-14
; NUMBER OF SEQ ID NOS: 115
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 34
; LENGTH: 42
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-09-930-512-34
```

```
Query Match 29.4%; Score 72; DB 11; Length 42;
Best Local Similarity 45.7%; Pred. No. 0.012;
Matches 16; Conservative 4; Mismatches 15; Indels 0; Gaps 0;
```

```
Qy 16 GLAESREQAKRLIMAGKVTLTNNSTTIPRLKPKG 50
Db 6 GFASSRSQARLIEHGHKVGKGVVTIPSYIVKPKG 40
```

```
RESULT 17
US-09-815-242-5030
; Sequence 5030, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
```

```
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 5030
; LENGTH: 201
; TYPE: PRT
; ORGANISM: Enterococcus faecalis
; US-09-815-242-5030
```

```
Query Match 29.0%; Score 71; DB 9; Length 201;
Best Local Similarity 34.7%; Pred. No. 0.12;
Matches 17; Conservative 8; Mismatches 16; Indels 8; Gaps 1;
```

```
Qy 3 KHKV-----RADELVFLQGLAESREQAKRLIMAGKVTLTNNSTTIP 43
Db 82 KHGVNFMVLEQRLDNVYRLGLATRRQARQLNVHGHITVDGKRVDP 130
```

```
RESULT 18
US-09-815-242-10879
; Sequence 10879, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 10879
; LENGTH: 203
; TYPE: PRT
; ORGANISM: Enterococcus faecalis
; US-09-815-242-10879
```

```
Query Match 29.0%; Score 71; DB 9; Length 203;
Best Local Similarity 34.7%; Pred. No. 0.12;
Matches 17; Conservative 8; Mismatches 16; Indels 8; Gaps 1;
```

```
Qy 3 KHKV-----RADELVFLQGLAESREQAKRLIMAGKVTLTNNSTTIP 43
Db 82 KHGVNFMVLEQRLDNVYRLGLATRRQARQLNVHGHITVDGKRVDP 130
```

```
RESULT 19
US-10-282-122A-42445
; Sequence 42445, Application US/10282122A
```

; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.

; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms

; FILE REFERENCE: ELITRA.034A

; CURRENT APPLICATION NUMBER: US/10/282,122A

; CURRENT FILING DATE: 2003-02-20

; PRIOR APPLICATION NUMBER: 60/191,078

; PRIOR FILING DATE: 2000-03-21

; PRIOR APPLICATION NUMBER: 60/206,848

; PRIOR FILING DATE: 2000-05-23

; PRIOR APPLICATION NUMBER: 60/207,727

; PRIOR FILING DATE: 2000-05-26

; PRIOR APPLICATION NUMBER: 60/230,335

; PRIOR FILING DATE: 2000-09-06

; PRIOR APPLICATION NUMBER: 60/230,347

; PRIOR FILING DATE: 2000-09-09

; PRIOR APPLICATION NUMBER: 60/242,578

; PRIOR FILING DATE: 2000-10-23

; PRIOR APPLICATION NUMBER: 60/253,625

; PRIOR FILING DATE: 2000-11-27

; PRIOR APPLICATION NUMBER: 60/257,931

; PRIOR FILING DATE: 2000-12-22

; PRIOR APPLICATION NUMBER: 60/267,636

; PRIOR FILING DATE: 2001-02-09

; PRIOR APPLICATION NUMBER: 60/269,308

; PRIOR FILING DATE: 2001-02-16

; Remaining Prior Application data removed - See File Wrapper or PALM.

; NUMBER OF SEQ ID NOS: 78614

; SOFTWARE: PatentIn version 3.1

; SEQ ID NO 42445

; LENGTH: 203

; TYPE: PRT

; ORGANISM: Enterococcus faecalis

; US-10-282-122A-42445

Query Match 29.0%; Score 71; DB 12; Length 203;

Best Local Similarity 34.7%; Pred. No. 0.12; Indels 8; Gaps 1;

Matches 17; Conservative 8; Mismatches 16; Indels 16; Gaps 1;

QY 3 KHKV-----RADELVFLQGLAESREQAKRLIMAGKVLTLNNSSTIP 43

Db 82 KKGVNPWVLEQLRDLNVVYRLGLATTRQARQLVNHGHITVDGKRYDIP 130

RESULT 20

US-09-738-626-4126

; Sequence 4126, Application US/09738626

; Publication No. US20020197605A1

; GENERAL INFORMATION:

; APPLICANT: NAKAGAWA, SATOSHI

; APPLICANT: MIZOGUCHI, HIROSHI

; APPLICANT: ANDO, SEIKO

; APPLICANT: HAYASHI, MIKIRO

; APPLICANT: OCHIAI, KEIKO

; APPLICANT: YOKOI, HARUHIKO

; APPLICANT: TATEISHI, NAOKO

; APPLICANT: SENOH, AKIHIRO

; APPLICANT: IKEDA, MASATO

; APPLICANT: OZAKI, AKIO

; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES

; FILE REFERENCE: 249-125

; CURRENT APPLICATION NUMBER: US/09/738,626

; CURRENT FILING DATE: 2000-12-18

; PRIOR APPLICATION NUMBER: JP 99/377484

; PRIOR FILING DATE: 1999-12-16

; PRIOR APPLICATION NUMBER: JP 00/159162

; PRIOR FILING DATE: 2000-04-07

; PRIOR APPLICATION NUMBER: JP 00/280988

; PRIOR FILING DATE: 2000-08-03

; NUMBER OF SEQ ID NOS: 7059

; SOFTWARE: PatentIn ver. 3.0

; SEQ ID NO 4126

; LENGTH: 201

; TYPE: PRT

; ORGANISM: Corynebacterium glutamicum

; US-09-738-626-4126

Query Match 28.6%; Score 70; DB 9; Length 201;

Best Local Similarity 35.1%; Pred. No. 0.17; Indels 15; Gaps 0;

Matches 13; Conservative 9; Mismatches 15; Indels 15; Gaps 0;

QY 7 RADELVFLQGLAESREQAKRLIMAGKVLTLNNSSTIP 43

Db 92 RLNDVVYRAGLANTRQARQLVSHGHFTVNGKADVP 128

RESULT 21

US-10-282-122A-53648

; Sequence 53648, Application US/10282122A

; Publication No. US20040029129A1

; GENERAL INFORMATION:

; APPLICANT: Wang, Liangsu

; APPLICANT: Zamudio, Carlos

; APPLICANT: Malone, Cheryl

; APPLICANT: Haselbeck, Robert

; APPLICANT: Ohlsen, Kari

; APPLICANT: Zyskind, Judith

; APPLICANT: Wall, Daniel

; APPLICANT: Trawick, John

; APPLICANT: Carr, Grant

; APPLICANT: Yamamoto, Robert

; APPLICANT: Forsyth, R.

; APPLICANT: Xu, H.

; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms

; FILE REFERENCE: ELITRA.034A

; CURRENT APPLICATION NUMBER: US/10/282,122A

; CURRENT FILING DATE: 2003-02-20

; PRIOR APPLICATION NUMBER: 60/191,078

; PRIOR FILING DATE: 2000-03-21

; PRIOR APPLICATION NUMBER: 60/206,848

; PRIOR FILING DATE: 2000-05-23

; PRIOR APPLICATION NUMBER: 60/207,727

; PRIOR FILING DATE: 2000-05-26

; PRIOR APPLICATION NUMBER: 60/230,335

; PRIOR FILING DATE: 2000-09-06

; PRIOR APPLICATION NUMBER: 60/230,347

; PRIOR FILING DATE: 2000-09-09

; PRIOR APPLICATION NUMBER: 60/242,578

; PRIOR FILING DATE: 2000-10-23

; PRIOR APPLICATION NUMBER: 60/253,625

; PRIOR FILING DATE: 2000-11-27

; PRIOR APPLICATION NUMBER: 60/257,931

; PRIOR FILING DATE: 2000-12-22

; PRIOR APPLICATION NUMBER: 60/267,636

; PRIOR FILING DATE: 2001-02-09

; PRIOR APPLICATION NUMBER: 60/269,308

; PRIOR FILING DATE: 2001-02-16

; Remaining Prior Application data removed - See File Wrapper or PALM.

; NUMBER OF SEQ ID NOS: 78614

; SOFTWARE: PatentIn version 3.1

; SEQ ID NO 53648

; LENGTH: 201

; TYPE: PRT

; ORGANISM: Corynebacterium diptheriae

US-10-282-122A-53648

Query Match 28.6%; Score 70; DB 12; Length 201;
Best Local Similarity 35.1%; Pred. No. 0.17;
Matches 13; Conservative 8; Mismatches 16; Indels

Qy 7 RADELVFLQGLAESREQAKRLIMAGKVTLTNNSTIP 43
| | : : | | : : | | : : | | : :
Db 92 RLDNVVYRAGLARTRRROAROLVSHGHFTVNGKKINVP 128

RESULT 22

```

US-10-369-493-22316
; Sequence 22316, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBI
; TITLE OF INVENTION: PLANTS WITH IMPROVED
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 22316
; LENGTH: 1147
; TYPE: PRT
; ORGANISM: Saccharomyces cerevisiae
US-10-369-493-22316

```

Query Match 28.2%; Score 69; DB 15; Length 1147;
Best Local Similarity 31.1%; Pred. No. 2.2;
Matches 14; Conservative 12; Mismatches 19; Indels

Qy 5 KVRADLVFLQGLAESREQAKRLIMACKVTLTNNSTTIPLRLEKP 49
 |:::| |:::| |:::| |:::| |:::| |:::| |:::|
D6 515 KLORRQALALQAQAQRQQQQOVALGTKVALLNNSPDIMTKMRSP 559

RESULT 23

```

US-10-618-581-2
; Sequence 2, Application US/10618581
; Publication NO. US20040077524A1
; GENERAL INFORMATION:
; APPLICANT: MANNING, GERARD
; APPLICANT: SUDARSANAM, SUCHA
; TITLE OF INVENTION: METHOD FOR SELECTIVELY INHIBITING FUNGAL GROWTH
; FILE REFERENCE: 034536/0323
; CURRENT APPLICATION NUMBER: US/10/618,581
; CURRENT FILING DATE: 2003-07-15
; PRIOR APPLICATION NUMBER: 60/395,624
; PRIOR FILING DATE: 2002-07-15
; NUMBER OF SEQ ID NOS: 42
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 1147
; TYPE: prt
; ORGANISM: Saccharomyces cerevisiae
US-10-618-581-2

```

```
Query Match      28.2%; Score 69; DB 16; Length 1147;
Best Local Similarity 31.1%; Pred. No. 2.2;
Matches 14; Conservative 12; Mismatches 19; Indels 0; Gaps 0;
```

QY 5 KVRDELVFLOGLAESREQAKRLIMACKVTLTNNSTTIPLRIEKP 49
 | : : : | : | : | : : : | : | : | : : | : : |
D6

D6 515 KLORRQALQAQAQRQQQQOVALGTKVALNNNSPDIMTRWESP 559

RESULT 24

```

US-10-282-122A-64188
; Sequence 64188, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Orlsen, Karl
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.

```

```

? TITLE: CAN. AD. RA.
? FIELD OF INVENTION: Identification of Essential Genes in Microorganisms
? FILE REFERENCE: ELITRA.034A
? CURRENT APPLICATION NUMBER: US/10/282,122A
? CURRENT FILING DATE: 2003-02-20
? PRIOR APPLICATION NUMBER: 60/191,078
? PRIOR FILING DATE: 2000-03-21
? PRIOR APPLICATION NUMBER: 60/206,848
? PRIOR FILING DATE: 2000-05-23
? PRIOR APPLICATION NUMBER: 60/207,727
? PRIOR FILING DATE: 2000-05-26
? PRIOR APPLICATION NUMBER: 60/230,335
? PRIOR FILING DATE: 2000-09-06
? PRIOR APPLICATION NUMBER: 60/230,347
? PRIOR FILING DATE: 2000-09-09
? PRIOR APPLICATION NUMBER: 60/242,578
? PRIOR FILING DATE: 2000-10-23
? PRIOR APPLICATION NUMBER: 60/253,625
? PRIOR FILING DATE: 2000-11-27
? PRIOR APPLICATION NUMBER: 60/257,931
? PRIOR FILING DATE: 2000-12-22
? PRIOR APPLICATION NUMBER: 60/267,636
? PRIOR FILING DATE: 2001-02-09
? PRIOR APPLICATION NUMBER: 60/269,308
? PRIOR FILING DATE: 2001-02-16
? Remaining Prior Application data removed - See File Wrapper or PALM.
? NUMBER OF SEQ ID NOS: 78614
? SOFTWARE: PatentIn version 3.1
? SEQ ID NO 64188
? LENGTH: 205
? TYPE: PRT
? ORGANISM: Mycoplasma pneumoniae
? US-10-282-122A-64188

```

Query Match 27.8%; Score 68; DB 12; Length 205;
Best Local Similarity 31.8%; Pred. No. 0.33;
Matches 14; Conservative 9; Mismatches 21; Indels

Qy 7 RADELVFLQLAESREQAKRLIMAGKVTLTNNSTTIPLRLEKPG 50
| | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Dd 96 RLDNIVYRMGEAPTFSARQLVNHHGVLLNDRTVDTPSIINPG 139

RESULT 25

```

US-10-282-122A-71298
; Sequence 71298, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haseibeck, Robert
; APPLICANT: Ohlsen, Karl
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John

```

```
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 71298
; LENGTH: 200
; TYPE: PRT
; ORGANISM: Staphylococcus haemolyticus
US-10-282-122A-71298

Query Match 27.3%; Score 67; DB 12; Length 200;
Best Local Similarity 37.2%; Pred. No. 0.45;
Matches 16; Conservative 7; Mismatches 20; Indels 0; Gaps 0;

QY 7 RADELVPLOGLAESRQAKRLIMAGKVTLTNNSTTIPLRLEKP 49
| | | | | | | | | | | | | | | | | | | | | |
Db 93 RLDAVYSLGLARTRQARQLVGHGHEVDGRRVDIPSYSLKP 135

RESULT 26
US-10-282-122A-48598
; Sequence 48598, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
```

```
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 48598
; LENGTH: 201
; TYPE: PRT
; ORGANISM: Bacteroides fragilis
US-10-282-122A-48598

Query Match 27.3%; Score 67; DB 12; Length 201;
Best Local Similarity 34.1%; Pred. No. 0.45;
Matches 15; Conservative 8; Mismatches 21; Indels 0; Gaps 0;

QY 7 RADELVPLOGLAESRQAKRLIMAGKVTLTNNSTTIPLRLEKPG 50
| | | | | | | | | | | | | | | | | | | | | |
Db 93 RLDNIVRLGIATPRAARQLVGHKHITVDGQVNVIFSYAVKPG 136

RESULT 27
US-10-282-122A-63563
; Sequence 63563, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
```

; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 63563
; LENGTH: 205
; TYPE: PRT
; ORGANISM: Mycoplasma genitalium
US-10-282-122A-63563

Query Match 27.3%; Score 67; DB 12; Length 205;
Best Local Similarity 29.5%; Pred. No. 0.46;
Matches 13; Conservative 11; Mismatches 20; Indels 0; Gaps 0;

QY 7 RADELVFQGLAESREQAKRLIMAGKVTLTNNSTTIPRLRLEKPG 50

DB 96 RLDNIVRMGFAPTRKRSARQWVNHGHVILNDQTVTPSIINPG 139

RESULT 28

US-10-282-122A-60489
; Sequence 60489, Application US/10282122A
; Publication No. US20040029129A1

; GENERAL INFORMATION:

; APPLICANT: Wang, Liangsu

; APPLICANT: Zamudio, Carlos

; APPLICANT: Malone, Cheryl

; APPLICANT: Haselbeck, Robert

; APPLICANT: Ohlsen, Kari

; APPLICANT: Zyskind, Judith

; APPLICANT: Wall, Daniel

; APPLICANT: Trawick, John

; APPLICANT: Carr, Grant

; APPLICANT: Yamamoto, Robert

; APPLICANT: Forsyth, R.

; APPLICANT: Xu, H.

; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms

; FILE REFERENCE: ELITRA.034A

; CURRENT APPLICATION NUMBER: US/10/282,122A

; CURRENT FILING DATE: 2003-02-20

; PRIOR APPLICATION NUMBER: 60/191,078

; PRIOR FILING DATE: 2000-03-21

; PRIOR APPLICATION NUMBER: 60/206,848

; PRIOR FILING DATE: 2000-05-23

; PRIOR APPLICATION NUMBER: 60/207,727

; PRIOR FILING DATE: 2000-05-26

; PRIOR APPLICATION NUMBER: 60/230,335

; PRIOR FILING DATE: 2000-09-06

; PRIOR APPLICATION NUMBER: 60/230,347

; PRIOR FILING DATE: 2000-09-09

; PRIOR APPLICATION NUMBER: 60/242,578

; PRIOR FILING DATE: 2000-10-23

; PRIOR APPLICATION NUMBER: 60/253,625

; PRIOR FILING DATE: 2000-11-27

; PRIOR APPLICATION NUMBER: 60/257,931

; PRIOR FILING DATE: 2000-12-22

; PRIOR APPLICATION NUMBER: 60/267,636

; PRIOR FILING DATE: 2001-02-09

; PRIOR APPLICATION NUMBER: 60/269,308

; PRIOR FILING DATE: 2001-02-16

; Remaining Prior Application data removed - See File Wrapper or PALM.

; NUMBER OF SEQ ID NOS: 78614

; SOFTWARE: PatentIn version 3.1

; SEQ ID NO 60489

; LENGTH: 200

; TYPE: PRT

; ORGANISM: Listeria monocytogenes

US-10-282-122A-60489

Query Match 26.9%; Score 66; DB 12; Length 200;

Best Local Similarity 35.1%; Pred. No. 0.62;

Matches 13; Conservative 8; Mismatches 16; Indels 0; Gaps 0;

QY 7 RADELVFQGLAESREQAKRLIMAGKVTLTNNSTTIP 43

DB 93 RLDNIVRGLGARTRAARQQLVNHGHITVDGKRVDP 129

RESULT 29

US-10-282-122A-76872

; Sequence 76872, Application US/10282122A

; Publication No. US20040029129A1

; GENERAL INFORMATION:

; APPLICANT: Wang, Liangsu

; APPLICANT: Zamudio, Carlos

; APPLICANT: Malone, Cheryl

; APPLICANT: Haselbeck, Robert

; APPLICANT: Ohlsen, Kari

; APPLICANT: Zyskind, Judith

; APPLICANT: Wall, Daniel

; APPLICANT: Trawick, John

; APPLICANT: Carr, Grant

; APPLICANT: Yamamoto, Robert

; APPLICANT: Forsyth, R.

; APPLICANT: Xu, H.

; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms

; FILE REFERENCE: ELITRA.034A

; CURRENT APPLICATION NUMBER: US/10/282,122A

; CURRENT FILING DATE: 2003-02-20

; PRIOR APPLICATION NUMBER: 60/191,078

; PRIOR FILING DATE: 2000-03-21

; PRIOR APPLICATION NUMBER: 60/206,848

; PRIOR FILING DATE: 2000-05-23

; PRIOR APPLICATION NUMBER: 60/207,727

; PRIOR FILING DATE: 2000-05-26

; PRIOR APPLICATION NUMBER: 60/230,335

; PRIOR FILING DATE: 2000-09-06

; PRIOR APPLICATION NUMBER: 60/230,347

; PRIOR FILING DATE: 2000-09-09

; PRIOR APPLICATION NUMBER: 60/242,578

; PRIOR FILING DATE: 2000-10-23

; PRIOR APPLICATION NUMBER: 60/253,625

; PRIOR FILING DATE: 2000-11-27

; PRIOR APPLICATION NUMBER: 60/257,931

; PRIOR FILING DATE: 2000-12-22

; PRIOR APPLICATION NUMBER: 60/267,636

; PRIOR FILING DATE: 2001-02-09

; PRIOR APPLICATION NUMBER: 60/269,308

; PRIOR FILING DATE: 2001-02-16

; Remaining Prior Application data removed - See File Wrapper or PALM.

; NUMBER OF SEQ ID NOS: 78614

; SOFTWARE: PatentIn version 3.1

; SEQ ID NO 76872

; LENGTH: 202

; TYPE: PRT

; ORGANISM: Ureaplasma urealyticum

US-10-282-122A-76872

Query Match 26.9%; Score 66; DB 12; Length 202;

Best Local Similarity 34.1%; Pred. No. 0.63;

Matches 15; Conservative 7; Mismatches 22; Indels 0; Gaps 0;

QY 7 RADELVFQGLAESREQAKRLIMAGKVTLTNNSTTIPRLRLEKPG 50

DB 95 RLDSIVYRAGFAPTRRAARQQLVNHSHVLVNNKKNIPSAIVEVG 138

RESULT 30

US-09-815-242-13240

; Sequence 13240, Application US/09815242

; Patent No. US20020061569A1

; GENERAL INFORMATION:

; APPLICANT: Haselbeck, Robert

; APPLICANT: Ohlsen, Kari L.

; APPLICANT: Zyskind, Judith W.

; APPLICANT: Wall, Daniel

; APPLICANT: Trawick, John D.

```

; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 73698
; LENGTH: 203
; TYPE: PRF
; ORGANISM: Streptococcus pneumoniae
US-10-282-122A-73698

Query Match                26.1%; Score 64; DB 12; Length 203;
Best Local Similarity      31.0%; Pred. No. 1.2;
Matches 14; Conservative   8; Mismatches 22; Indels 0; Gaps 0

Qy    7 RAEELVFLQGAAESRQAOKLIMAGVKVLTNNSTTTPRLKKG 50
       |||::|||::|::|::|::|::|::|::|::|::|::|::|
Db    94 RLDNVVYRLGATTRQARQGVNHHGHILVDGRKDIPSYRVTEG 137

RESULT 32
US-10-282-122A-74814
; Sequence 74814, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganism
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 74814
; LENGTH: 203
; TYPE: PRF
; ORGANISM: Streptococcus pyogenes
```

```
US-10-282-122A-74814
Query Match          26.1%; Score 64; DB 12; Length 203;
Best Local Similarity 31.8%; Pred. No. 1.2; Indels 0; Gaps 0;
Matches 14; Conservative 8; Mismatches 22; Indels 0; Gaps 0;

Qy 7 RADELVFQGLAESREQAQKRLIMAGKVTLTNNSTTIFRLKPG 50
| | | | | | | | | | | | | | | | | | | | | |
Db 94 RLDNVVYRLGLATTRQARQFVNHGHILVDGKRVDPISYRVDPG 137

RESULT 33
US-10-474-776-751
; Sequence 751, Application US/10474776
; Publication No. US20040110181A1
; GENERAL INFORMATION:
; APPLICANT: Wyeth
; TITLE OF INVENTION: NOVEL STREPTOCOCCUS PNEUMONIAE OPEN READING FRAMES ENCODING POLYPEPTIDES
; FILE REFERENCE: AM100649-PCT
; CURRENT APPLICATION NUMBER: US/10/474,776
; CURRENT FILING DATE: 2003-10-14
; NUMBER OF SEQ ID NOS: 752
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 751
; LENGTH: 203
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-10-474-776-751

Query Match          26.1%; Score 64; DB 16; Length 203;
Best Local Similarity 31.8%; Pred. No. 1.2; Indels 0; Gaps 0;
Matches 14; Conservative 8; Mismatches 22; Indels 0; Gaps 0;

Qy 7 RADELVFQGLAESREQAQKRLIMAGKVTLTNNSTTIFRLKPG 50
| | | | | | | | | | | | | | | | | | | | | |
Db 94 RLDNVVYRLGLATTRQARQFVNHGHILVDGKRVDPISYRVTPG 137

RESULT 34
US-10-282-122A-47209
; Sequence 47209, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
```

```
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 47209
; LENGTH: 209
; TYPE: PRT
; ORGANISM: Borrelia burgdorferi
US-10-282-122A-47209

Query Match          25.7%; Score 63; DB 12; Length 209;
Best Local Similarity 35.1%; Pred. No. 1.7; Indels 0; Gaps 0;
Matches 13; Conservative 7; Mismatches 17; Indels 0; Gaps 0;

Qy 7 RADELVFQGLAESREQAQKRLIMAGKVTLTNNSTTIP 43
| | | | | | | | | | | | | | | | | | | | | |
Db 97 RLDNVVYRAGFAISRARQIVSHGIIILNGRRVTIP 133

RESULT 35
US-10-335-977-7740
; Sequence 7740, Application US/10335977
; Publication No. US20040052799A1
; GENERAL INFORMATION:
; APPLICANT: DOUGLAS SMITH et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO HELICOBACTER PYLORI FOR DIAGNOSTICS AND THERAPEUTICS
; NUMBER OF SEQUENCES: 10031
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD
; STREET: 28 State Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02109-1875
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD-ROM ISO9660
; COMPUTER: IBM PC Compatible
; OPERATING SYSTEM: Windows NT 4.0
; SOFTWARE: UNIX
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/335,977
; FILING DATE: 30-Dec-2002
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/993,002
; FILING DATE: 17-DEC-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Mandragouras, Amy E.
; REGISTRATION NUMBER: 36,207
; REFERENCE/DOCKET NUMBER: GTN-018
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)227-7400
; TELEFAX: (617)742-4214
; INFORMATION FOR SEQ ID NO: 7740:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 235 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHEICAL: YES
; ORIGINAL SOURCE:
; ORGANISM: Helicobacter pylori
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (B) LOCATION 1...235
; SEQUENCE DESCRIPTION: SEQ ID NO: 7740:
```

```

; Publication No. US20040052799A1
; GENERAL INFORMATION:
; APPLICANT: DOUGLAS SMITH et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES
; RELATING TO HELICOBACTER PYLORI FOR
; DIAGNOSTICS AND THERAPEUTICS
;
; NUMBER OF SEQUENCES: 10031
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD
; STREET: 28 State Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02109-1875
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD-ROM ISO9660
; COMPUTER: IBM PC Compatible
; OPERATING SYSTEM: Windows NT 4.0
; SOFTWARE: UNIX
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: 08/993,002
; FILING DATE: 17-DEC-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/10/335,977
; FILING DATE: 30-Dec-2002
; ATTORNEY/AGENT INFORMATION:
; NAME: Mandragouras, Amy E.
; REGISTRATION NUMBER: 36,207
; REFERENCE/DOCKET NUMBER: GTN-018
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 227-7400
; TELEFAX: (617) 742-4214
; INFORMATION FOR SEQ ID NO: 7742:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 246 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: YES
; ORIGINAL SOURCE:
; ORGANISM: Helicobacter pylori
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (B) LOCATION 1...246
; SEQUENCE DESCRIPTION: SEQ ID NO: 7742:
;
US-10-335-977-7742
;
Query Match          25.7%; Score 63; DB 12; Length 246;
Best Local Similarity 35.6%; Pred.No. 2.2;
Matches 16; Conservative 9; Mismatches 20; Indels 0; Gaps 0;

QY      6 VRADELVFLQGIAESREQAKRLIMACKGVLTNNSTIPLRLEKPG 50
       :| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:|
DB      12 MRLDYALFNQHLSANSEKAKLVLLKKQVLNVKNWVSKPSFIYKEG 56

RESULT 38
US-10-282-122A-61700
; Sequence 61700, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.

```



```
/ TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
/ FILE REFERENCE: ELITRA.034A
/ CURRENT APPLICATION NUMBER: US/10/282,122A
/ CURRENT FILING DATE: 2003-02-20
/ PRIOR APPLICATION NUMBER: 60/191,078
/ PRIOR FILING DATE: 2000-03-21
/ PRIOR APPLICATION NUMBER: 60/206,848
/ PRIOR FILING DATE: 2000-05-23
/ PRIOR APPLICATION NUMBER: 60/207,727
/ PRIOR FILING DATE: 2000-05-26
/ PRIOR APPLICATION NUMBER: 60/230,335
/ PRIOR FILING DATE: 2000-09-06
/ PRIOR APPLICATION NUMBER: 60/230,347
/ PRIOR FILING DATE: 2000-09-09
/ PRIOR APPLICATION NUMBER: 60/242,578
/ PRIOR FILING DATE: 2000-10-23
/ PRIOR APPLICATION NUMBER: 60/253,625
/ PRIOR FILING DATE: 2000-11-27
/ PRIOR APPLICATION NUMBER: 60/257,931
/ PRIOR FILING DATE: 2000-12-22
/ PRIOR APPLICATION NUMBER: 60/267,636
/ PRIOR FILING DATE: 2001-02-09
/ PRIOR APPLICATION NUMBER: 60/269,308
/ PRIOR FILING DATE: 2001-02-16
/ Remaining Prior Application data removed - See File Wrapper or PALM.
/ NUMBER OF SEQ ID NOS: 78614
/ SOFTWARE: PatentIn version 3.1
/ SEQ ID NO 61700
/ LENGTH: 201
/ TYPE: PRT
/ ORGANISM: Mycobacterium avium
US-10-282-122A-61700

Query Match      24.9%; Score 61; DB 12; Length 201;
Best Local Similarity 32.4%; Pred. No. 3.2;
Matches 12; Conservative 8; Mismatches 17; Indels 0; Gaps 0;

Qy      7 RADELVFQGLAESREQAKRLIMAGKVTLTNNSTTIP 43
Db      92 RLDNVVYRAGLARTRRMARQLVTHGHTVNGVRDVP 128

RESULT 39
US-10-282-122A-72484
/ Sequence 72484, Application US/10282122A
/ Publication No. US20040029129A1
/ GENERAL INFORMATION:
/ APPLICANT: Wang, Liangsu
/ APPLICANT: Zamudio, Carlos
/ APPLICANT: Malone, Cheryl
/ APPLICANT: Haselbeck, Robert
/ APPLICANT: Ohlsen, Kari
/ APPLICANT: Zyskind, Judith
/ APPLICANT: Wall, Daniel
/ APPLICANT: Trawick, John
/ APPLICANT: Carr, Grant
/ APPLICANT: Yamamoto, Robert
/ APPLICANT: Forsyth, R.
/ APPLICANT: Xu, H.
/ TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
/ FILE REFERENCE: ELITRA.034A
/ CURRENT APPLICATION NUMBER: US/10/282,122A
/ CURRENT FILING DATE: 2003-02-20
/ PRIOR APPLICATION NUMBER: 60/191,078
/ PRIOR FILING DATE: 2000-03-21
/ PRIOR APPLICATION NUMBER: 60/206,848
/ PRIOR FILING DATE: 2000-05-23
/ PRIOR APPLICATION NUMBER: 60/207,727
/ PRIOR FILING DATE: 2000-05-26
/ PRIOR APPLICATION NUMBER: 60/230,335
/ PRIOR FILING DATE: 2000-09-06
/ PRIOR APPLICATION NUMBER: 60/230,347
/ PRIOR FILING DATE: 2000-09-09
/ PRIOR APPLICATION NUMBER: 60/242,578
/ PRIOR FILING DATE: 2000-10-23
/ PRIOR APPLICATION NUMBER: 60/253,625
/ PRIOR FILING DATE: 2000-11-27
/ PRIOR APPLICATION NUMBER: 60/257,931
/ PRIOR FILING DATE: 2000-12-22
/ PRIOR APPLICATION NUMBER: 60/267,636
/ PRIOR FILING DATE: 2001-02-09
/ PRIOR APPLICATION NUMBER: 60/269,308
/ PRIOR FILING DATE: 2001-02-16
/ Remaining Prior Application data removed - See File Wrapper or PALM.
/ NUMBER OF SEQ ID NOS: 78614
/ SOFTWARE: PatentIn version 3.1
/ SEQ ID NO 72484
/ LENGTH: 203
/ TYPE: PRT
/ ORGANISM: Streptococcus mutans
US-10-282-122A-72484

Query Match      24.5%; Score 60; DB 12; Length 203;
Best Local Similarity 32.4%; Pred. No. 4.5;
Matches 12; Conservative 8; Mismatches 17; Indels 0; Gaps 0;

Qy      7 RADELVFQGLAESREQAKRLIMAGKVTLTNNSTTIP 43
Db      94 RLDNVVYRAGLARTRRMARQLVTHGHTVNGVRDVP 130

RESULT 40
US-10-282-122A-74807
/ Sequence 74807, Application US/10282122A
/ Publication No. US20040029129A1
/ GENERAL INFORMATION:
/ APPLICANT: Wang, Liangsu
/ APPLICANT: Zamudio, Carlos
/ APPLICANT: Malone, Cheryl
/ APPLICANT: Haselbeck, Robert
/ APPLICANT: Ohlsen, Kari
/ APPLICANT: Zyskind, Judith
/ APPLICANT: Wall, Daniel
/ APPLICANT: Trawick, John
/ APPLICANT: Carr, Grant
/ APPLICANT: Yamamoto, Robert
/ APPLICANT: Forsyth, R.
/ APPLICANT: Xu, H.
/ TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
/ FILE REFERENCE: ELITRA.034A
/ CURRENT APPLICATION NUMBER: US/10/282,122A
/ CURRENT FILING DATE: 2003-02-20
/ PRIOR APPLICATION NUMBER: 60/191,078
/ PRIOR FILING DATE: 2000-03-21
/ PRIOR APPLICATION NUMBER: 60/206,848
/ PRIOR FILING DATE: 2000-05-23
/ PRIOR APPLICATION NUMBER: 60/207,727
/ PRIOR FILING DATE: 2000-05-26
/ PRIOR APPLICATION NUMBER: 60/230,335
/ PRIOR FILING DATE: 2000-09-06
/ PRIOR APPLICATION NUMBER: 60/230,347
/ PRIOR FILING DATE: 2000-09-09
/ PRIOR APPLICATION NUMBER: 60/242,578
/ PRIOR FILING DATE: 2000-10-23
/ PRIOR APPLICATION NUMBER: 60/253,625
/ PRIOR FILING DATE: 2000-11-27
/ PRIOR APPLICATION NUMBER: 60/257,931
/ PRIOR FILING DATE: 2000-12-22
/ PRIOR APPLICATION NUMBER: 60/267,636
/ PRIOR FILING DATE: 2001-02-09
/ PRIOR APPLICATION NUMBER: 60/269,308
/ PRIOR FILING DATE: 2001-02-16
/ Remaining Prior Application data removed - See File Wrapper or PALM.
/ NUMBER OF SEQ ID NOS: 78614
/ SOFTWARE: PatentIn version 3.1
/ SEQ ID NO 74807
```


GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 7, 2004, 16:52:00 ; Search time 4.48505 Seconds
(without alignments)
1072.357 Million cell updates/sec

Title: US-10-009-919A-1_COPY_1_50

Perfect score: 245

Sequence: 1 MAKHKVRADELVFLQGLAES.....GKVLTTNNSTTIPRLRKPG 50

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR 78.*

1: pir1.*

2: pir2.*

3: pir3.*

4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	99	40.4	269	2 G69967	hemolysin homolog
2	96	39.2	271	2 B98007	conserved hypotet
3	90	36.7	267	2 A72338	hemolysin - Thermo
4	88	35.9	274	2 AB1508	conserved hypotet
5	88	35.9	274	2 AF1245	conserved hypotet
6	82	33.5	200	1 A37146	ribosomal protein
7	81.5	33.3	272	2 B83997	hemolysin-like pro
8	80	32.7	200	2 A84051	ribosomal protein
9	80	32.7	267	2 H97155	probable rRNA meth
10	78	31.8	211	2 A70307	ribosomal protein
11	77.5	31.6	274	2 S76154	hypothetical prote
12	75.5	30.8	268	2 S72968	cytotoxin/hemolysi
13	75.5	30.8	269	2 H87078	cytotoxin/hemolysi
14	75	30.6	209	2 JE0399	30S ribosomal prot
15	74	30.2	200	2 C89566	serine/threonine-s
16	69	28.2	1147	2 S64330	30S ribosomal prot
17	68	27.8	199	2 D90601	30S ribosomal prot
18	68	27.8	203	2 D86660	30S ribosomal prot
19	68	27.8	205	2 S73721	30S ribosomal prot
20	67	27.3	205	2 D64234	ribosomal protein
21	66.5	27.1	268	2 E70502	probable tlyA prot
22	66	26.9	187	2 G64323	ribosomal protein
23	66	26.9	200	2 AD1274	ribosomal protein
24	66	26.9	200	2 AE1637	ribosomal protein
25	66	26.9	201	2 T06900	ribosomal protein
26	66	26.9	202	2 A82883	ribosomal protein
27	66	26.9	209	2 B72247	ribosomal protein
28	66	26.9	258	2 H70367	hemolysin - Aquife
29	64	26.1	203	2 H95009	ribosomal protein

30	64	26.1	203	2 F97881	30S ribosomal prot
31	64	26.1	257	2 S55251	ribosomal protein
32	64	26.1	271	2 T36885	probable membrane
33	63	25.7	209	2 F70176	ribosomal protein
34	63	25.7	235	2 F71944	probable hemolysin
35	62.5	25.5	257	2 C86784	pseudouridine synt
36	62.5	25.5	304	2 D81182	adhesin, probable
37	62	25.3	206	2 S78281	ribosomal protein
38	61	24.9	240	2 R43863	cytotoxin/hemolysi
39	60	24.5	200	1 R3MD4	ribosomal protein
40	60	24.5	202	1 R3LV4	ribosomal protein
41	60	24.5	202	2 S74386	ribosomal protein
42	60	24.5	463	2 D81960	ATP-dependent DNA
43	59.5	24.3	169	1 T43938	ribosomal protein
44	59.5	24.3	170	2 B84259	30S ribosomal prot
45	59	24.1	160	2 S47021	ribosomal protein

ALIGNMENTS

RESULT 1

G69967

hemolysin homolog yqxG - Bacillus subtilis

N;Alternate names: hypothetical protein 65 (ahrC 5' region)

C;Species: Bacillus subtilis

C;Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 20-Jun-2000

C;Accession: G69967; PS0053

R;Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Berter C.; Bron, S.; Brouillet, S.; Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Cho A.; Ehrlich, S.D.; Emmerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E. Nature 390, 249-256, 1997

A;Authors: Foulger, D.; Fritz, C.; Fujita, Y.; Funa, S.; Galizzi, A.; Gallert, Jech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F.; Koeter, P.; Konigstein, G.; Krogh, M.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois, A.;Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maueel, Y, M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelle, Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadaie, Y.; Sato, T.; Scanlon, A;Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Seror, akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiyama, T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K. A;Authors: Yoshikawa, H.F.; Zumstein, E.; Yoshikawa, H.; Danchin, A. A;Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis. A;Reference number: A69580; MUID:98044033; PMID:9384377

A;Accession: G69967

A;Status: nucleic acid sequence not shown; translation not shown

A;Molecule type: DNA

A;Residues: 1-269 <KUN>

A;Cross-references: GB:Z99116; GB:AL009126; NID:G2634723; PIDN:CAB14357.1; PID:G2634860

A;Experimental source: strain 168

R;North, A.K.; Smith, M.C.M.; Baumberg, S.

Gene 80, 29-38, 1989

A;Title: Nucleotide sequence of a Bacillus subtilis arginine regulatory gene and homolog

A;Reference number: A91616; MUID:90006783; PMID:2507400

A;Accession: PS0053

A;Molecule type: DNA

A;Residues: 178-179, 'LIAQ', 211-269 <NOR>

A;Cross-references: GB:M27869; NID:G142448; PIDN:AAA22207.1; PID:G142449

A;Experimental source: strain EMG50

C;Genetics:

A;Gene: yqxG

C;Superfamily: hemolysin homolog yqxG

Query Match 40.4%; Score 99; DB 2; Length 269;

Best Local Similarity 50.0%; Pred. No. 5e-05;

Matches 25; Conservative 5; Mismatches 14; Indels 6; Gaps 1;

QY 1 MAKHKVRADELVFLQGLAESRQAKRLIMAGKVLTNNSTTIPRLRKPG 50

Db 1 MTSKKERLDVLLVERGLAETREKAKRAIMAGIVYSNEN-----RIDKPG 44

RESULT 2

[illegible]


```

A:Accession: S72968
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-268 <SMI>
A:Cross-references: EMBL:U00021; NID:g467141; PIDN:AAA50906.1; PID:g467142
C:Genetics:
C:Start codon: GTG
C:Superfamily: hemolysin homolog yqxX

Query Match          30.8%; Score 75.5; DB 2; Length 268;
Best Local Similarity 51.4%; Pred. No. 0.053;
Matches 18; Conservative 8; Mismatches 8; Indels 1; Gaps

QY      1 MAKHKVRADLVFLQGLAESREQAKLIMAGKVTL 35
       |||::|||::|||::|||::|||::|||::|||::
DB      1 MAR-RVRVDVELVRRGLARSQQAAKLIISAGKSVI 34

RESULT 13
H87078 cytoxin/hemolysin [imported] - Mycobacterium leprae
C:Species: Mycobacterium leprae
C>Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 10-May-2000
C:Accession: H87078
R:Coile, S.F.; Eiglmier, K.; Parkhill, J.; James, K.D.; Thomson, N.R.; Wheeler, R.; Davies, R.M.; Devlin, K.; Duthoy, S.; Felwell, T.; Fraser, A.; Hamlin, M.A.; Rutherford, K.M.
Nature 409, 1007-1011, 2001
A:Authors: Rutter, S.; Seeger, K.; Simon, S.; Simmonds, M.; Skelton, J.; Smeeth, P.L.
A>Title: Massive gene decay in the leprosy bacillus.
A:Reference number: A86909, PMID:21128732; PMID:11234002
A:Accession: H87078
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-269 <STO>
A:Cross-references: GB:AL450380; NID:gi3093256; PIDN:CAC31739.1; GSPDB:XNO
C:Genetics:
A:Gene: tlyA
C:Superfamily: hemolysin homolog yqxX

Query Match          30.8%; Score 75.5; DB 2; Length 269;
Best Local Similarity 51.4%; Pred. No. 0.053;
Matches 18; Conservative 8; Mismatches 8; Indels 1; Gaps

QY      1 MAKHKVRADLVFLQGLAESREQAKLIMAGKVTL 35
       |||::|||::|||::|||::|||::|||::|||::
DB      1 MAR-RVRVDVELVRRGLARSQQAAKLIISAGKSVI 34

RESULT 14
JE0399 ribosomal protein S4 - Thermus aquaticus (strain HB8)
C:Species: Thermus aquaticus
C>Date: 23-Jul-1999 #sequence_revision 23-Jul-1999 #text_change 11-May-2000
C:Accession: JE0399
R:Wada, T.; Yamazaki, T.; Kuramitsu, S.; Kyogoku, Y.
J. Biochem 125, 143-150, 1999
A>Title: Cloning of the RNA polymerase alpha subunit gene from Thermus thermophilus.
A:Reference number: JE0397; PMID:99098837; PMID:9980810
A:Accession: JE0399
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-209 <WAD>
C:Genetics:
A:Gene: rps4
C:Superfamily: Escherichia coli ribosomal protein S4

Query Match          30.6%; Score 75; DB 2; Length 209;
Best Local Similarity 34.1%; Pred. No. 0.047;
Matches 15; Conservative 9; Mismatches 20; Indels 0; Gaps

QY      7 RADELVFLOGLAESREQAKLIMAGKVTLNNSTTIFLRLEKPG 50
       |||::|||::|||::|||::|||::|||::|||::

```


A:Gene: rpsD
C:Superfamily: Escherichia coli ribosomal protein S4

Query Match 26.9%; Score 66; DB 2; Length 200;
Best Local Similarity 35.1%; Pred. No. 0.64;
Matches 13; Conservative 8; Mismatches 16; Indels 0; Gaps 0;

Qy 7 RADELVFQGLAESREQAKRLIMAGKVTLTNNSTTIP 43
| | | | | | | | | | | | | | | | | | | | | |
Db 93 RLDNIVYRLGLARTERRAARQLVNHGHITVDGKRVDP 129

RESULT 24

AE1637
ribosomal protein S4 [imported] - Listeria innocua (strain Clip11262)
C:Species: Listeria innocua
C:Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 14-Dec-2001
C:Accession: AB1637
R:Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloeker
.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi, H.
D.; Jones, L.M.; Karst, U.
Science 294, 849-852, 2001
A:Authors: Kretz, J.; Kuhn, M.; Kunst, F.; Kurapat, G.; Madueno, E.; Maitournam, A.; Ma
ok, C.; Schlueter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehland,
A.; Title: Comparative genomics of Listeria species.
A:Reference number: AB1077; MUID:21537279; PMID:11679669
A:Accession: AB1637
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-200 <GLA>
A:Cross-references: GB:AL592022; PIDN:CAC96869.1; PID:g16414125; GSPDB:GN00178
A:Experimental source: strain Clip11262
C:Genetics:
A:Gene: rpsD
C:Superfamily: Escherichia coli ribosomal protein S4

Query Match 26.9%; Score 66; DB 2; Length 200;
Best Local Similarity 35.1%; Pred. No. 0.64;
Matches 13; Conservative 8; Mismatches 16; Indels 0; Gaps 0;

Qy 7 RADELVFQGLAESREQAKRLIMAGKVTLTNNSTTIP 43
| | | | | | | | | | | | | | | | | | | | | |
Db 93 RLDNIVYRLGLARTERRAARQLVNHGHITVDGKRVDP 129

RESULT 25

T06900
ribosomal protein S4 - Cyanophora paradoxa cyanelle
C:Species: cyanelle Cyanophora paradoxa
C:Date: 30-Apr-1999 #sequence_revision 30-Apr-1999 #text_change 13-Aug-1999
C:Accession: T06900
R:Stirewalt, V.L.; Michalowski, C.B.; Luffelhardt, W.; Bohnert, H.J.; Bryant, D.A.
submitted to the EMBL Data Library, July 1995
A:Description: Nucleotide sequence of the cyanelle genome from Cyanophora paradoxa.
A:Reference number: Z15840
A:Accession: T06900
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-201 <STI>
A:Cross-references: EMBL:U30821; NID:g1016083; PIDN:AAA81243.1; PID:g1016156
A:Experimental source: strain Pringsheim LB555
C:Genetics:
A:Gene: rps4
A:Genome: cyanelle
C:Superfamily: Escherichia coli ribosomal protein S4
C:Keywords: cyanelle; ribosome

Query Match 26.9%; Score 66; DB 2; Length 201;
Best Local Similarity 30.4%; Pred. No. 0.64;
Matches 14; Conservative 12; Mismatches 20; Indels 0; Gaps 0;

Qy 5 KVADELIVFQGLAESREQAKRLIMAGKVTLTNNSTTIPRLRLEKPG 50
: | | | | | | | | | | | | | | | | | | | | | |

Db 90 EMRLDNIVFRLGMAPTIPASRLQVNHGHICVNNKVVSIPIYQCKPG 135

RESULT 26

A82883
ribosomal protein S4 UU498 [imported] - Ureaplasma urealyticum
C:Species: Ureaplasma urealyticum
C:Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 02-Sep-2000
C:Accession: A82883
R:Glass, J.I.; Lefkowitz, E.J.; Glass, J.S.; Heiner, C.R.; Chen, E.Y.; Cassell, G.H.
submitted to GenBank, February 2000
A:Description: The complete sequence of Ureaplasma urealyticum: Alternate views of a min
A:Reference number: A82870
A:Accession: A82883
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-202 <GLA>
A:Cross-references: GB:AF002148; GB:AF222894; NID:g68999495; PIDN:AAF30910.1; GSPDB:GN001;
A:Experimental source: serovar 3; biovar 1
C:Genetics:
A:Gene: rps4; UU498
A:Genetic code: SGC3
C:Superfamily: Escherichia coli ribosomal protein S4

Query Match 26.9%; Score 66; DB 2; Length 202;
Best Local Similarity 34.1%; Pred. No. 0.65;
Matches 15; Conservative 7; Mismatches 22; Indels 0; Gaps 0;

Qy 7 RADELVFQGLAESREQAKRLIMAGKVTLTNNSTTIPRLRLEKPG 50
| | | | | | | | | | | | | | | | | | | | | |
Db 95 RLDNLVYRAGFAPTRRAARQLVNHSHLVNKNKVNIPISALVEVG 138

RESULT 27

B72247
ribosomal protein S4 - Thermotoga maritima (strain MSB8)
C:Species: Thermotoga maritima
C:Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 21-Jul-2000
C:Accession: B72247
R:Nelson, K.E.; Clayton, R.A.; Gill, S.R.; Gwinn, M.L.; Dodson, R.J.; Haft, D.H.; Hickey,
Garrett, M.M.; Stewart, A.M.; Cotton, M.D.; Pratt, M.S.; Phillips, C.A.; Richardson, D.;
C.M.
Nature 399, 323-329, 1999
A:Title: Evidence for lateral gene transfer between Archaea and Bacteria from genome seq
A:Reference number: A72200; MUID:99287316; PMID:10360571
A:Accession: B72247
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-209 <ARN>
A:Cross-references: GB:AE001798; GB:AE000512; NID:g4982033; PIDN:AAD36541.1; PID:g498203;
C:Genetics:
A:Gene: TMI473
C:Superfamily: Escherichia coli ribosomal protein S4

Query Match 26.9%; Score 66; DB 2; Length 209;
Best Local Similarity 31.1%; Pred. No. 0.67;
Matches 14; Conservative 9; Mismatches 22; Indels 0; Gaps 0;

Qy 5 KVADELIVFQGLAESREQAKRLIMAGKVTLTNNSTTIPRLRLEKPG 49
: | | | | | | | | | | | | | | | | | | | | | |
Db 97 EARLDNVVYRMGFAINRQARQLVNHGHFLVNGKKNIPSYLLRP 141

RESULT 28

H70367
hemolysin - Aquifex aeolicus
C:Species: Aquifex aeolicus
C:Date: 08-May-1998 #sequence_revision 08-May-1998 #text_change 24-Sep-1999
C:Accession: H70367
R:Deckert, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, D.E.; Ove
V.
Nature 392, 353-358, 1998

A;Title: The complete genome of the hyperthermophilic bacterium Aquifex aeolicus.

A;Reference number: A70300; MUID:98196666; PMID:9537320

A;Accession: H70367

A;Status: preliminary; nucleic acid sequence not shown; translation not shown

A;Molecule type: DNA

A;Residues: 1-258 <AQF>

A;Cross-references: GB:AE000707; NID:g2983342; PIDN:AAC06935.1; PID:g2983347; GB:AE00065

A;Experimental source: strain VF5

C;Genetics:

A;Gene: tly

C;Superfamily: hemolysin homolog yqx

Query Match 26.9%; Score 66; DB 2; Length 258;

Best Local Similarity 37.8%; Pred. No. 0.84;

Matches 17; Conservative 11; Mismatches 11; Indels 6; Gaps 2;

QY 6 VRADELVFLQGLAESRQAKRLIMAGKVTLTNSTTIPRLKPG 50

DB 1 MRLDKYLTDKGIVPSREKAQAVIMAGQV-LVNGKV-----VDKPG 39

RESULT 29

H95009 ribosomal protein S4 [imported] - Streptococcus pneumoniae (strain TIGR4)

C;Species: Streptococcus pneumoniae

C;Date: 03-Aug-2001 #sequence_revision 03-Aug-2001 #text_change 24-Aug-2001

C;Accession: H95009

R;Tetelin, H.; Nelson, K.E.; Paulsen, I.T.; Eisen, J.A.; Read, T.D.; Peterson, S.; Held

on, J.D.; Umayam, L.A.; White, O.; Salzberg, S.L.; Lewis, M.R.; Radune, D.; Holtzapple,

nson, T.; Hickey, E.K.; Holt, I.E.

Science 293, 498-506, 2001

A;Authors: Loftus, B.J.; Yang, F.; Smith, H.O.; Venter, J.C.; Dougherty, B.A.; Morrison,

A;Title: Complete Genome Sequence of a virulent isolate of Streptococcus pneumoniae.

A;Reference number: A95000; MUID:21357209; PMID:11463316

A;Accession: H95009

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-203 <KUR>

A;Cross-references: GB:AE005672; PIDN:AAK74273.1; PID:g14971552; GSPDB:GN00164; TIGR:SP4

A;Experimental source: strain TIGR4

C;Genetics:

A;Gene: SP0085

C;Superfamily: Escherichia coli ribosomal protein S4

Query Match 26.1%; Score 64; DB 2; Length 203;

Best Local Similarity 31.8%; Pred. No. 1.2;

Matches 14; Conservative 8; Mismatches 22; Indels 0; Gaps 0;

QY 7 RADELVFLQGLAESRQAKRLIMAGKVTLTNSTTIPRLKPG 50

DB 94 RLDNVVYRLGLATTRQARQFVNHGHILVDGKRVDPSPYRVTPG 137

RESULT 30

F97881

30S ribosomal protein S4 [imported] - Streptococcus pneumoniae (strain R6)

C;Species: Streptococcus pneumoniae

C;Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 02-Nov-2001

C;Accession: F97881

R;Hoskins, J.A.; Alborn Jr., W.; Arnold, J.; Blaszcak, L.; Burgett, S.; DeHoff, B.S.; E

e, R.; LeBlanc, D.J.; Lee, L.N.; Lefkowitz, E.J.; Lu, J.; Matsushima, P.; McAhren, S.; M

Y, P.; Sun, P.M.; Winkler, M.E.

J. Bacteriol. 183, 5709-5717, 2001

A;Authors: Yang, Y.; Young-Bellido, M.; Zhao, G.; Zook, C.; Baltz, R.H.; Jaskunas, S.R.;

A;Title: Genome of the Bacterium Streptococcus pneumoniae Strain R6.

A;Reference number: A97872; MUID:21429245; PMID:11544234

A;Accession: F97881

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-203 <KUR>

A;Cross-references: GB:AE007317; PIDN:AAK98882.1; PID:g15457613; GSPDB:GN00174

C;Genetics:

A;Gene: rpsD

C;Superfamily: Escherichia coli ribosomal protein S4

Query Match 26.1%; Score 64; DB 2; Length 203;

Best Local Similarity 31.8%; Pred. No. 1.2;

Matches 14; Conservative 8; Mismatches 22; Indels 0; Gaps 0;

QY 7 RADELVFLQGLAESRQAKRLIMAGKVTLTNSTTIPRLKPG 50

DB 94 RLDNVVYRLGLATTRQARQFVNHGHILVDGKRVDPSPYRVTPG 137

RESULT 31

S5251

ribosomal protein S4, chloroplast - Chlamydomonas reinhardtii chloroplast

C;Species: chloroplast Chlamydomonas reinhardtii

C;Date: 23-Aug-1995 #sequence_revision 19-Oct-1995 #text_change 13-Aug-1999

C;Accession: S5251

R;Randolph-Anderson, B.L.; Boynton, J.E.; Gillham, N.W.; Huang, C.; Liu, X.Q.

Mol. Gen. Genet. 247, 295-305, 1995

A;Title: The chloroplast gene encoding ribosomal protein S4 in Chlamydomonas reinhardtii

utation in ribosomal protein S12.

A;Reference number: S5251; MUID:95287849; PMID:7770034

A;Accession: S5251

A;Molecule type: DNA

A;Residues: 1-257 <RAN>

A;Cross-references: EMBL:U17357; NID:g862378; PIDN:AAA81363.1; PID:g862379

C;Genetics:

A;Gene: rps4

A;Genome: chloroplast

C;Superfamily: Escherichia coli ribosomal protein S4

C;Keywords: chloroplast; protein biosynthesis; ribosome

Query Match 26.1%; Score 64; DB 2; Length 257;

Best Local Similarity 31.1%; Pred. No. 1.5;

Matches 14; Conservative 10; Mismatches 21; Indels 0; Gaps 0;

QY 5 KVRADSLVFLQGLAESRQAKRLIMAGKVTLTNSTTIPRLKPG 49

DB 109 EMRLDNIVFLNMAPTIPARQLISHGRVNNKKNIPSYMCKP 153

RESULT 32

T36885

probable membrane protein - Streptomyces coelicolor

C;Species: Streptomyces coelicolor

C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 11-Jan-2000

C;Accession: T36885

R;Murphy, L.; Harris, D.; James, K.D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.

submitted to the EMBL Data Library, August 1999

A;Reference number: Z21617

A;Accession: T36885

A;Status: preliminary; translated from GB/EMBL/DBJ

A;Molecule type: DNA

A;Residues: 1-271 <MUR>

A;Cross-references: EMBL:AL109848; PIDN:CAB52846.1; GSPDB:GN00070; SCOPDB:SC151.22c

A;Experimental source: strain A3(2)

C;Genetics:

A;Gene: SCOPDB:SC151.22c

C;Superfamily: hemolysin homolog yqx

Query Match 26.1%; Score 64; DB 2; Length 271;

Best Local Similarity 43.2%; Pred. No. 1.6;

Matches 16; Conservative 5; Mismatches 16; Indels 0; Gaps 0;

QY 7 RADELVFLQGLAESRQAKRLIMAGKVTLTNSTTIP 43

DB 8 RLDAELVRRKLARSREHASQLAAGRVTVGKTVATKP 44

RESULT 33

F70176

ribosomal protein S4 (rpsD) - Lyme disease spirochete

C;Species: Borrelia burgdorferi (Lyme disease spirochete)

C>Date: 13-Feb-1998 #sequence_revision 13-Feb-1998 #text_change 13-Aug-1999
C/Accession: F70176

R/Fraser, C.M.; Casjens, S.; Huang, W.M.; Sutton, G.G.; Clayton, R.; Lathigra, R.; Whitson, D.; Peterson, J.; Kervilavage, A.R.; Ouackenbush, J.; Salzberg, S.; Hanson, M.; Vugt, B.; Bowman, C.; Garland, S.; Fujii, C.; Cotton, M.D.; Horst, K.; Roberts, K.; Hatch, B.

Nature 390, 580-586, 1997

A;Authors: Smith, H.O.; Venter, J.C.
A>Title: Genomic sequence of a Lyme disease spirochaete, Borrelia burgdorferi.
A/Reference number: A70100; UID:98065943; PMID:9403685
A/Accession: F70176
A>Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-209 <KLE>
A/Cross-references: GB:AE001162; GB:AE000783; NID:g2688528; PIDN:AAC66961.1; PID:g2688528
A/Experimental source: strain B31
C/Superfamily: Escherichia coli ribosomal protein S4

Query Match 25.7%; Score 63; DB 2; Length 209;
Best Local Similarity 35.1%; Pred.No.1.6;
Matches 13; Conservative 7; Mismatches 17; Indels 0; Gaps 0;

Qy 7 RADELVFLGGLAESREQAQLIMAGKVLTNNSTTP 43
Db 97 RIDNVVRGAFAISRAHQIVSHGIILNGRRVTIP 133

| | : | : | | | : | | : | : | : |
| | : | : | | | : | | : | : | : |

RESULT 34
F71944
probable hemolysin - Helicobacter pylori (strain J99)
C/Species: Helicobacter pylori
A/Variety: strain J99
C/Date: 12-Feb-1999 #sequence_revision 12-Feb-1999 #text_change 24-Sep-1999
R/Alm, R.A.; Ling, L.S.L.; Moir, D.T.; King, B.L.; Brown, E.D.; Doig, P.C.; Smith, D.R.; Ives, C.; Gibson, R.; Werberg, D.; Mills, S.D.; Jiang, Q.; Taylor, D.E.; Vovis, G.F.; Nature 397, 176-180, 1999
A>Title: Genomic sequence comparison of two unrelated isolates of the human gastric pathogen Helicobacter pylori.
A/Reference number: A71800; UID:99120557; PMID:9923682
A/Accession: F71944
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-235 <AKN>
A/Cross-references: GB:AE001469; GB:AE001439; NID:g4154861; PIDN:AAO5917.1; PID:g4154861
A/Experimental source: strain J99
C/Genetics:
A/Gene: hlyA
C/Superfamily: hemolysin homolog yqxC

Query Match 25.7%; Score 63; DB 2; Length 235;
Best Local Similarity 35.6%; Pred.No.1.8;
Matches 16; Conservative 9; Mismatches 20; Indels 0; Gaps 0;

Qy 6 VRADLVFLGGLAESREQAQLIMAGKVLTNNSTTPIRLKP 50
Db 1 MLRDYALFNQHLSAREKAALVLKKQVLNMWVKSPFIIVEK 45

: | : | : | | | : | : | : | : | : |
: | : | : | | | : | : | : | : | :

RESULT 35
C86784
Pseudouridine synthase [imported] - Lactococcus lactis subsp. lactis (strain IL1403)
C/Species: Lactococcus lactis subsp. lactis
C/Date: 23-Mar-2001 #sequence_revision 23-Mar-2001 #text_change 03-Aug-2001
C/Accession: C86784
R/Bolorin, A.; Wincker, P.; Mauger, S.; Jaillon, O.; Malarme, K.; Weissenbach, J.; Ehrlich Genome Res. 11, 731-753, 2001
A>Title: The complete genome sequence of the lactic acid bacterium Lactococcus lactis ss
A/Reference number: AB6625; UID:21235186; PMID:11337471
A/Accession: C86784
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-257 <STO>
A/Cross-references: GB:AE005176; PID:gl2724251; PIDN:AAK05373.1; GSDB:GN00146
A/Experimental source: strain IL1403

Mon Jul 12 15:30:40 2004

Db 94 EMRLDTICTLGFAPTIAARQLVNHGHITVNDNVVSIIP 132

RESULT 38

A3863

Cytotoxin/hemolysin ORF2 tly - Treponema hyodysenteriae

C:Species: Treponema hyodysenteriae

C>Date: 04-Sep-1998 #sequence_revision 04-Sep-1998 #text_change 26-Aug-1999

C:Accession: A43863

R:Muir, S.; Koopman, M.B.; Libby, S.J.; Joens, L.A.; Heffron, F.; Kusters, J.G.

Infect. Immun. 60, 529-535, 1992

A:Title: Cloning and expression of a Serpula (Treponema) hyodysenteriae hemolysin gene.

A:Reference number: A43863; MUID:92112315; PMID:1730486

A:Accession: A43863

A>Status: Preliminary

A:Molecule type: DNA

A:Residues: 1-240 <MUI>

A:Experimental source: strain B204

A>Note: sequence extracted from NCBI backbone (NCBIN:77501, NCBIIP:77504)

C:Superfamily: hemolysin homolog ykx

Query Match 24.9%; Score 61; DB 2; Length 240;

Best Local Similarity 38.9%; Pred. No. 3.4; Mismatches 15; Indels 0; Gaps 0;

Matches 14; Conservative 7;

QY 6 VRADELVFLQGLAESREQAKRLIMAGKVTLTNSTT 41

Db 1 MRLDEVHSEGYTESRSKAQDILLACGVFVGKVT 36

RESULT 39

R3MD4

ribosomal protein S4, chloroplast - Cryptomonas sp. chloroplast (strain phi)

C:Species: chloroplast Cryptomonas sp.

A:Variety: strain phi

C>Date: 31-Dec-1990 #sequence_revision 31-Dec-1990 #text_change 30-Jan-1998

C:Accession: S09219

R:Douglas, S.E.; Durnford, D.G.

Nucleic Acids Res. 18, 1903, 1990

A:Title: Nucleotide sequence of the genes for ribosomal protein S4 and tRNA-Arg from the

A:Reference number: S09219; MUID:90245597; PMID:2336372

A:Accession: S09219

A:Molecule type: DNA

A:Residues: 1-200 <DOU>

A:Cross-references: EMBL:X51511; NID:g18103; PID:g18104

C:Genetics:

A:Gene: rps4

A:Genome: chloroplast

C:Superfamily: Escherichia coli ribosomal protein S4

C:Keywords: chloroplast; protein biosynthesis; ribosome

Query Match 24.5%; Score 60; DB 1; Length 200;

Best Local Similarity 28.2%; Pred. No. 3.8; Mismatches 16; Indels 0; Gaps 0;

Matches 11; Conservative 12;

QY 5 KVRDELVFLQGLAESREQAKRLIMAGKVTLTNSTTIP 43

Db 89 EMRLDNVIFRLGWAPTIPAAQQLVNHGHIKVNNTRVSIIP 127

RESULT 40

R3LV4

ribosomal protein S4, chloroplast - liverwort (Marchantia polymorpha) chloroplast

C:Species: chloroplast Marchantia polymorpha

C>Date: 30-Jun-1987 #sequence_revision 30-Jun-1987 #text_change 22-Jun-1999

C:Accession: A02704; S01602

R:Ohyama, K.

submitted to the EMBL Data Library, October 1986

A:Reference number: A00150

A:Accession: A02704

A:Molecule type: DNA

A:Residues: 1-202 <OHY>

R:Ohyama, K.; Fukuzawa, H.; Kohchi, T.; Shirai, H.; Sano, T.; Sano, S.; Umesono, K.; Shi

Nature 322, 572-574, 1986

A:Title: Chloroplast gene organization deduced from complete sequence of liverwort March

A:Reference number: A38014

R:Umesono, K.; Inokuchi, H.; Shiki, Y.; Takeuchi, M.; Chang, Z.; Fukuzawa, H.; Kohchi, T

J. Mol. Biol. 203, 299-331, 1988

A:Title: Structure and organization of Marchantia polymorpha chloroplast genome. II. Gen

A:Reference number: S01567; MUID:89068686; PMID:2974085

A:Accession: S01602

A:Molecule type: DNA

A:Residues: 1-202 <UME>

A:Cross-references: GB:X04465; GB:Y00686; NID:g11640; PIDN:CAA28086.1; PID:g11674

C:Genetics:

A:Gene: rps4

A:Genome: chloroplast

C:Superfamily: Escherichia coli ribosomal protein S4

C:Keywords: chloroplast; protein biosynthesis; ribosome

Query Match 24.5%; Score 60; DB 1; Length 202;

Best Local Similarity 28.9%; Pred. No. 3.8; Mismatches 20; Indels 0; Gaps 0;

Matches 13; Conservative 12;

QY 5 KVRDELVFLQGLAESREQAKRLIMAGKVTLTNSTTIPRLRLEKP 49

Db 89 EMRLDNVIFRLGWAPTIPGARGLVNHRHILINNNTVDIPSYNCKP 133

Search completed: July 7, 2004, 16:57:35

Job time : 4.48505 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 7, 2004, 16:47:59 ; Search time 3.32226 Seconds
(without alignments)
783.656 Million cell updates/sec

Title: US-10-009-919A-1_COPY_1_50

Perfect score: 245

Sequence: 1 MAKHKVRADELVFLQGLAES.....GKVLTNNSTTIPLRLEKPG 50

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_42:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	99	40.4	269	1 YQXC BACSU	P19672 bacillus su
2	85	34.7	200	1 RS4_OCEIH	P59130 oceanobacil
3	82	33.5	199	1 RS4_BACSU	P21466 bacillus su
4	81	33.1	199	1 RS4_BACST	P81288 bacillus st
5	80	32.7	200	1 RS4_BACHD	O9K728 bacillus ha
6	80	32.7	206	1 RS4_THETN	O8r7Y1 thermoanaer
7	78	31.8	211	1 RS4_AQUAE	O66484 aquifex aeo
8	75	30.6	208	1 RS4_THETH	P80373 thermus the
9	74	30.2	200	1 RS4_STAAM	Q997e4 staphylococ
10	73	29.8	200	1 RS4_BACAA	Q81kt2 bacillus an
11	73	29.8	200	1 RS4_BACCR	Q817a4 bacillus ce
12	73	29.8	200	1 RS4_STAEP	O8cs54 staphylococ
13	72	29.4	202	1 RS4_LACPL	Q88ux0 lactobacill
14	70	28.6	201	1 RS4_CORGL	Q8nsv4 corynebacte
15	69	28.2	200	1 RS4_THEVO	Q97b95 thermoplas
16	69	28.2	202	1 RS4_TAKLE	P59150 takakia lep
17	69	28.2	1147	1 KIN2 YEAST	P31186 saccharomyc
18	68	27.8	199	1 RS4_MVCPU	Q98pk6 mycoplasma
19	68	27.8	203	1 RS4_LACLA	Q8cis2 lactococcus
20	68	27.8	205	1 RS4_MYCPN	P46775 mycoplasma
21	67	27.3	205	1 RS4_MYCYE	P47553 mycoplasma
22	66	26.9	187	1 RS4_METJA	P54020 methanococc
23	66	26.9	200	1 RS4_LISIN	Q92bb2 listeria in
24	66	26.9	200	1 RS4_LISMO	O8v6t6 listeria mo
25	66	26.9	201	1 RS4_CYAPA	P48133 cyanophora
26	66	26.9	202	1 RS4_UREPA	Q9ppz2 ureaplasma
27	66	26.9	205	1 RS4_THEMA	Q9xli3 thermotoga
28	65	26.5	209	1 RS4_MESVI	Q9mun0 mesostigma
29	64	26.1	203	1 RS4_STRPN	Q97f69 streptococc
30	64	26.1	203	1 RS4_STRPY	Q99xj4 streptococc
31	64	26.1	257	1 RS4_CHLRE	P48270 chlamydomon
32	63	25.7	203	1 RS4_CHLTE	P59129 chlorobium
33	63	25.7	209	1 RS4_BORBU	O51560 borrelia bu

RESULT 1

ID	YQXC BACSU	STANDARD;	PRT;	269 AA.
AC	P19672;			
DT	01-FEB-1991 (Rel. 17, Created)			
DT	01-OCT-1996 (Rel. 34, Last sequence update)			
DT	10-OCT-2003 (Rel. 42, Last annotation update)			
DE	Hypothetical protein YQXC.			
GN	YQXC OR YQIF OR BSU24260.			
OS	Bacillus subtilis.			
OC	Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.			
OX	NCBI_TaxID=1423;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=168 / JH642;			
RX	MEDLINE=97124195; PubMed=8969508;			
RA	Mizuno M., Masuda S., Takemaru K.-I., Hosono S., Sato T., Takeuchi M., Kobayashi Y.;			
RT	"Systematic sequencing of the 283 kb 210 degrees-232 degrees region of the Bacillus subtilis genome containing the skin element and many sporulation genes."			
RL	Microbiology 142:3103-3111(1996).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=168;			
RX	MEDLINE=98044033; PubMed=9384377;			
RA	Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G., Szabo A., Borsari R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S., Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M., Choi S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A., Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T., Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D., Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N., Guiseppe G., Guy B.-J., Haga K., Haiech J., Harwood C.R., Henaut A., Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L., Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C., Kobayashi Y., Koetter P., Koningsstein G., Krogh S., Kumano M., Kurita K., Lapius A., Lardinois S., Lauber J., Lazarevic V., Lee S.M., Levine A., Liu H., Masuda S., Maestl D., Nakai S., Noback M., Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H., Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M., Presecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S., Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadie Y., Sato T., Scanlan E., Schleich S., Schroeter R., Scofield F., Sekiguchi J., Sekowska A., Seror S.J., Seror P., Shin B.S., Soldo B., Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K., Takeuchi M., Takakoshi A., Tanaka T., Terstra P., Tognoni A., Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A., Viari A., Wambut R., Wedler E., Wedler H., Weitzenecker T., Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K., Yoshida K., Yoshikawa H.F., Zumstein E., Yoshikawa H., Danchin A.;			
RT	"The complete genome sequence of the Gram-positive bacterium Bacillus subtilis.";			

ALIGNMENTS

34	62	25.3	206	1	RR4_ODOSI	P49492 odontella s
35	61	24.9	199	1	RS4_THEAC	O9hjd7 thermoplas
36	61	24.9	240	1	HLA_TREHY	Q06803 treponema h
37	60	24.5	200	1	RR4_GUITH	P17072 guillardia
38	60	24.5	202	1	RR4_MARPO	P06358 marchantia
39	60	24.5	202	1	RR4_PLARP	Q9m4c3 plagiochasm
40	60	24.5	202	1	RS4_SYNEL	P59134 synechococc
41	60	24.5	202	1	RS4_SYNY3	P48939 synechocyst
42	60	24.5	203	1	RS4_STRMU	P59133 streptococc
43	59.5	24.3	169	1	RS4_HALHA	Q9v2w3 halobacteri
44	59.5	24.3	170	1	RS4_HALNI	Q9hjq6 halobacteri
45	59.5	24.3	582	1	SYD_STRP3	Q8x5j0 streptococc

```
RL Nature 390:249-256(1997).
RN [3]
RP SEQUENCE OF 205-269 FROM N.A.
RC STRAIN=168 / EMG50;
RA MEDLINE=90006783; PubMed=2507400;
RX North A.K., Smith M.C.M., Baumberg S.;
RT "Nucleotide sequence of a Bacillus subtilis arginine regulatory gene
RT and homology of its product to the Escherichia coli arginine
RT repressor."
RL Gene 80:29-38(1989).
CC -!- SIMILARITY: Contains 1 S4 RNA-binding domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; D84432; BAA12577.1; -
CC EMBL; Z99116; CAB14357.1; -
CC EMBL; M27869; AAA22207.1; -
CC PIR; G69967; G69967.
CC Subtilisin; BG10308; YQXC.
CC InterPro; IPR002877; RrmJ_FtsJ.
CC InterPro; IPR002942; S4.
CC InterPro; IPR004538; Tly.
CC Pfam; PF01728; FtsJ; 1.
CC Pfam; PF01479; S4; 1.
CC SMART; SM00363; S4; 1.
CC TIGRPFAMs; TIGR00478; tly; 1.
CC PROSITE; PS50889; S4; 1.
CC Hypothetical protein; RNA-binding; Complete proteome.
KW FT DOMAIN 6 67
FT CONFLICT 205 210 HADVLK -> CMLIAQ (IN REF. 3).
FT SEQUENCE 269 AA; 29652 MW; 77802BC3F8AC97F6 CRC64;
SQ
Query Match 40.4%; Score 99; DB 1; Length 269;
Best Local Similarity 50.08; Pred. No. 4.6e-05;
Matches 25; Conservative 5; Mismatches 14; Indels 6; Gaps 1;
Oy 1 MAKHKVRADELVFLQGLAEAREQAKELIMAGKVTLTNNSTTIPRLKPG 50
| | | | | : | | | | | : | | | | | : | | | | |
Db 1 MTSKRLDLVLVERGLAETREKAKRAIMAGIVYSNEN-----RLDKPG 44
| | | | | : | | | | | : | | | | | : | | | | |
RESULT 2
RS4_OCEIH STANDARD; PRT; 200 AA.
AC PS9130;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE 30S ribosomal protein S4.
GN RPSD OR OB2205.
OS Oceanobacillus iheyensis.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Oceanobacillus.
OX NCBI_TaxID=182710;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=HTE831 / DSM 14371 / JCM 11309;
RX MEDLINE=2220767; PubMed=12235376;
RA Takami H., Takaki Y., Uchiyama I.;
RT "Genome sequence of Oceanobacillus iheyensis isolated from the Iheya
RT Ridge and its unexpected adaptive capabilities to extreme
RT environments."
RL Nucleic Acids Res. 30:3927-3935(2002).
CC -!- FUNCTION: One of the primary rRNA binding proteins, it binds
CC directly to 16S rRNA where it nucleates assembly of the body of
CC the 30S subunit (By similarity).
CC -!- FUNCTION: With S5 and S12 plays an important role in translational
CC accuracy (By similarity).
```

```
CC -!- SUBUNIT: Part of the 30S ribosomal subunit. Contacts protein S5.
CC The interaction surface between S4 and S5 is involved in control
CC of translational fidelity (By similarity).
CC -!- SIMILARITY: Contains 1 S4 RNA-binding domain.
CC -!- SIMILARITY: Belongs to the S4P family of ribosomal proteins.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AP004600; BAC14161.1; -
CC HAMAP; MF_01306; -; 1.
CC InterPro; IPR001912; Ribosomal_S4.
CC InterPro; IPR005709; Ribosomal_S4_b/o.
CC InterPro; IPR002942; S4.
CC Pfam; PF00163; Ribosomal_S4; 1.
CC Pfam; PF01479; S4; 1.
CC SMART; SM00363; S4; 1.
CC TIGRPFAMs; TIGR01017; rpsD_bact; 1.
CC PROSITE; PS00632; RIBOSOMAL_S4; 1.
CC PROSITE; PS50889; S4; 1.
CC Ribosomal protein; RNA-binding; rRNA-binding; Complete proteome.
KW FT DOMAIN 92 152 S4 RNA-BINDING.
FT CONFLICT 200 AA; 23040 MW; D64B3F3E497CFFC8 CRC64;
SQ
Query Match 34.7%; Score 85; DB 1; Length 200;
Best Local Similarity 43.2%; Pred. No. 0.0019;
Matches 19; Conservative 6; Mismatches 19; Indels 0; Gaps 0;
Oy 7 RADELVFLQGLAEAREQAKELIMAGKVTLTNNSTTIPRLKPG 50
| | | | | : | | | | | : | | | | | : | | | | |
Db 93 RLNLVYRLGLARTRQARQLVNHGHTVDGGRVDIPSYAIKPG 136
| | | | | : | | | | | : | | | | | : | | | | |
RESULT 3
RS4_BACSU STANDARD; PRT; 199 AA.
ID RS4_BACSU STANDARD; PRT; 199 AA.
AC P21466; O45662; Q53282;
DT 01-MAY-1991 (Rel. 18, Created)
DT 01-AUG-1991 (Rel. 19, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE 30S ribosomal protein S4 (BS4).
GN RPSD OR BSU29660.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=98048467; PubMed=9387221;
RA Lapidus A., Galleron N., Sorokin A., Ehrlich S.D.;
RT "Sequencing and functional annotation of the Bacillus subtilis genes
RT in the 200 kb rnbB-dnaB region."
RL Microbiology 143:3431-3441(1997).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=98044033; PubMed=9384377;
RA Kunst F., Ogasawara N., Moser J., Albertini A.M., Alloni G.,
RA Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,
RA Borriss R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,
```

RA Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,
RA Choi S.K., Codani J.J., Conneron I.F., Cummings N.J., Daniel R.A.,
RA Denicot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T.,
RA Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,
RA Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
RA Ghim S.Y., Glaser P., Goffeau A., Gollighly E.J., Grandi G.,
RA Guiseppi G., Guy B.J., Haga K., Halech J., Harwood C.R., Henaut A.,
RA Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
RA Joris B., Karamata D., Kasahara Y., Klaer-Blanchard M., Klein C.,
RA Kobayashi Y., Koetter P., Koningsstein G., Krogh S., Kumano M.,
RA Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
RA Lee S.M., Levine A., Liu H., Masuda S., Maui C., Medigue C.,
RA Medina N., Mellado R.P., Mizuno M., Moesti D., Nakai S., Noback M.,
RA Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,
RA Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,
RA Presecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
RA Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadaie Y.,
RA Sato T., Scallan E., Schleich S., Schroeter R., Scoffone F.,
RA Seguchi J., Sekowska A., Seror S.J., Serror P., Shin B.S., Soldo B.,
RA Sorokin A., Tacconi E., Takagi T., Takahashi H., Takenaru K.,
RA Takeuchi M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni A.,
RA Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,
RA Viari A., Wambit R., Wedler E., Wedler H., Weitzenegger T.,
RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
RA Yoshida K., Yoshikawa H.F., Zumsstein E., Yoshikawa H., Danchin A.,
RT "The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis*.";
RL Nature 390:249-256(1997).
RN [4]
RN SEQUENCE OF 1-56 FROM N.A.
RX MEDLINE=91310564; PubMed=1906866;
RP Grundy F.J., Henkin T.M.;
RA "The rpsD gene, encoding ribosomal protein S4, is autogenously
RT regulated in *Bacillus subtilis*.";
RN J. Bacteriol. 173:4595-4602(1991).
RN [5]
RN SEQUENCE OF 1-24.
RX MEDLINE=92219212; PubMed=6806564;
RA Higo K.-I., Otake E., Osawa S.;
RT "Purification and characterization of 30S ribosomal proteins from
RT *Bacillus subtilis*: correlation to *Escherichia coli* 30S proteins.";
RL Mol. Gen. Genet. 185:239-244(1982).
RN [6]
RN VARIANTS.
RX MEDLINE=91035249; PubMed=2121712;
RA Henkin T.M., Chambliss G.H., Grundy F.;
RT "Bacillus subtilis mutants with alterations in ribosomal protein S4.";
RL J. Bacteriol. 172:6380-6385(1990).
RN [7]
RN CHARACTERIZATION, AND VARIANTS.
RC STRAIN=168;
RX MEDLINE=21382165; PubMed=11489846;
RA Inaoka T., Kasai K., Ochi K.;
RT "Construction of an in vivo nonsense readthrough assay system and
RT functional analysis of ribosomal proteins S12, S4, and S5 in *Bacillus subtilis*.";
RL J. Bacteriol. 183:4958-4963(2001).
CC -I- FUNCTION: One of the primary rRNA binding proteins, it binds
CC directly to 16S rRNA where it nucleates assembly of the body of
CC the 30S subunit (By similarity).
CC -I- FUNCTION: With S5 and S12 plays an important role in translational
CC accuracy; many suppressors of streptomycin-dependent mutants of
CC protein S12 are found in this protein, some but not all of which
CC decrease translational accuracy (ram, ribosomal ambiguity
CC mutations).
CC -I- FUNCTION: S4 represses its own expression; it is not known if this
CC is at the level of translation or of mRNA stability.
CC -I- SUBUNIT: Part of the 30S ribosomal subunit. Contacts protein S5.
CC The interaction surface between S4 and S5 is involved in control
CC of translational fidelity.
CC -I- SIMILARITY: Contains 1 S4 RNA-binding domain.
CC -I- SIMILARITY: Belongs to the S4P family of ribosomal proteins.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M59358; AAA22717.1; -;
DR EMBL; AF008220; AAC00397.1; -;
DR EMBL; S45404; AAB19387.1; -;
DR EMBL; Z99119; CAB14944.1; -;
DR EMBL; M60889; AAA22716.1; -;
DR PIR; A37146; A37146.
DR HSSP; P81288; IC05.
DR Subtilist; BGI0372; rpsD.
DR HAMAP; MF_01306; -; 1.
DR InterPro; IPR001912; Ribosomal_S4.
DR InterPro; IPR005709; Ribosomal_S4_b/o.
DR InterPro; IPR002942; S4.
DR Pfam; PF00163; Ribosomal_S4; 1.
DR Pfam; PF01479; S4; 1.
DR SMART; SM00363; S4; 1.
DR TIGRfams; TIGR01017; rpsD_bact; 1.
DR PROSITE; PS00632; RIBOSOMAL_S4; 1.
DR PROSITE; PS00889; S4; 1.
KW Repressor; Ribosomal protein; RNA-binding; rRNA-binding;
KW Complete proteome.
FT INIT MET 0
FT DOMAIN 91 154 S4 RNA-BINDING.
FT VARIANT 45 45 E -> K (IN RPSD3; SUPPRESSES S12
FT MUTATION K55D).
FT VARIANT 74 77 MISSING (IN RPSD2; SUPPRESSES S12
FT MUTATION K55D. A RAM MUTATION).
FT VARIANT 77 77 L -> LAGKL (IN RPSD1; SUPPRESSES S12
FT MUTATION K55D. A RAM MUTATION).
SQ SEQUENCE 199 AA; 22704 MW; 23D669C542E22134 CRC64;
Query Match 33.5%; Score 82; DB 1; Length 199;
Best Local Similarity 38.6%; Pred. No. 0.0046;
Matches 17; Conservative 9; Mismatches 18; Indels 0; Gaps 0;
Qy 7 RADELVELQGLAESREQAKRLIMAGKVTLTNSTTIPRLKPG 50
Db 92 RLDDNVYKLGARTROARQVNHGHLVDGSRVDIFSILVKPG 135
RESULT 4
RS4_BACST
ID RS4_BACST STANDARD; PRT; 199 AA.
AC P81288;
DT 15-DEC-1998 (Rel. 37, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE 30S ribosomal protein S4 (BS5).
GN RPSD.
OS *Bacillus stearothermophilus*.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; *Geobacillus*.
OC NCBI_TaxID=1422;
RN [1]
RP SEQUENCE.
RX MEDLINE=92110431; PubMed=1764513;
RA Arndt E., Scholzen T., Kromer W., Hatakeyama T., Kimura M.;
RT "Primary structures of ribosomal proteins from the archaeobacterium
RT *Halobacterium marismortui* and the eubacterium *Bacillus stearothermophilus*.";
RT Biochimie 73:657-668(1991).
RN [2]
RP SEQUENCE FROM N.A., AND X-RAY CRYSTALLOGRAPHY (1.7 ANGSTROMS) OF
RP 41-199.
RX MEDLINE=98372721; PubMed=9707415;
RA Davies C., Gerstner R.B., Draper D.E., Ramakrishnan V., White S.W.;
RT "The crystal structure of ribosomal protein S4 reveals a two-domain

Mon

	SEQUENCE	211 AA;	24815 MW;	D9567456A594E1B8 CRC64;	
	Query Match	31.8%;	Score 78;	DB 1; Length 211;	
	Best Local Similarity	36.4%;	Pred. No. 0.015;		
	Matches	16;	Conservative	8; Mismatches 20; Indels 0; Gaps 0;	
Oy	7 RADELVFLQGLAESBEQAQRLLIMAGKVTILNNSITIPLRLEKPG	50			
Db	99 RLDDNVYVLFGASTRQARQLVAHGVLVNGKKVNPISYLVEPG	142			
	RESULT 8				
ID_	RS4_THETH	STANDARD;	PRT;	208 AA.	
AC	P80373; Q9Z9H7;				
DT	01-FEB-1995 (Rel. 31, Created)				
DD	30-MAY-2000 (Rel. 39, Last sequence update)				
DE	10-OCT-2003 (Rel. 42, Last annotation update)				
DE	30S ribosomal protein S4.				
GN	RPSD OR RPS4.				
OS	Thermus thermophilus.				
OC	Bacteria; Deinococcus-Thermus; Deinococci; Thermales; Thermaceae;				
OX	Thermus.				
NCBI_TaxID=	274;				
[1]	_TaxID=274;				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=HB8 / ATCC 27634;				
RC	MEDLINE=99098837; PubMed=9880810;				
RA	Wada T., Yamazaki T., Kuramitsu S., Kyogoku Y.;				
RT	"Cloning of the RNA polymerase alpha subunit gene from Thermus				
RT	thermophilus HB8 and characterization of the protein.";				
J. Biochem.	125:143-150(1999).				
[2]					
RP	SEQUENCE OF 1-26.				
RC	STRAIN=HB8 / ATCC 27634;				
RC	MEDLINE=95045586; PubMed=7957245;				
RA	Tsiboli P., Herfurth E., Choli T.;				
RT	"Purification and characterization of the 30S ribosomal proteins from				
RT	the bacterium Thermus thermophilus.";				
Eur. J. Biochem.	226:169-177(1994).				
[3]					
RP	X-RAY CRYSTALLOGRAPHY (5.5 ANGSTROMS) OF THE 30S SUBUNIT.				
RC	MEDLINE=99404610; PubMed=10476960;				
RA	Clemons W.M. Jr., May J.L.C., Wimberly B.T., McCutcheon J.P.,				
RA	Capel M.S., Ramakrishnan V.;				
RT	"Structure of a bacterial 30S ribosomal subunit at 5.5 A resolution.";				
Nature	400:833-840(1999).				
[4]					
RP	X-RAY CRYSTALLOGRAPHY (3.05 ANGSTROMS) OF THE 30S SUBUNIT.				
RC	MEDLINE=20466110; PubMed=11014182;				
RA	Wimberly B.T., Brodersen D.E., Clemons W.M. Jr., Morgan-Warren R.J.,				
RA	Carter A.P., Vonrhein C., Hartsch T., Ramakrishnan V.;				
RT	"Structure of the 30S ribosomal subunit.";				
Nature	407:327-339(2000).				
[5]					
RP	X-RAY CRYSTALLOGRAPHY (3.3 ANGSTROMS) OF THE 30S SUBUNIT.				
RC	MEDLINE=20460546; PubMed=11007480;				
RA	Schluenzen F., Tocilj A., Zarivach R., Harms J., Gluehmann M.,				
RA	Janelli D., Bashan A., Bartels H., Agmon I., Franceschi F., Yonath A.;				
RT	"Structure of functionally activated small ribosomal subunit at 3.3				
A resolution".					
Cell	102:615-623(2000).				
[6]					
RP	X-RAY CRYSTALLOGRAPHY (3.3 ANGSTROMS) OF THE 30S SUBUNIT.				
RC	MEDLINE=21097713; PubMed=11163189;				
RA	Brodersen D.E., Clemons W.M. Jr., Carter A.P., Morgan-Warren R.J.,				
RA	Wimberly B.T., Ramakrishnan V.;				
RT	"The structural basis for the action of the antibiotics tetracycline,				
pactamycin, and hygromycin B on the 30S ribosomal subunit.";					
Cell	103:1143-1154(2000).				
[7]					
RP	X-RAY CRYSTALLOGRAPHY (3.00 ANGSTROMS) OF THE 30S SUBUNIT.				
RC	MEDLINE=20466111; PubMed=11014183;				
RA	Carter A.P., Clemons W.M. Jr., Brodersen D.E., Morgan-Warren R.J.,				
RA	Wimberly B.T., Ramakrishnan V.;				
RT	"Functional insights from the structure of the 30S ribosomal subunit				
and its interactions with antibiotics.";					
Nature	407:340-348(2000).				
[8]					
RP	X-RAY CRYSTALLOGRAPHY (5.00 ANGSTROMS) OF THE RIBOSOME.				
RC	MEDLINE=21402420; PubMed=11511350;				
RA	Yusupova G.Z., Yusupov M.M., Cate J.H.D., Noller H.F.;				
RT	"The path of messenger RNA through the ribosome.";				
Cell	106:233-241(2001).				
[9]					
RP	X-RAY CRYSTALLOGRAPHY (3.2 ANGSTROMS) OF THE 30S SUBUNIT.				
RC	STRAIN=HB27;				
RC	MEDLINE=21192023; PubMed=11296217;				
RA	Pioletti M., Schluenzen F., Harms J., Zarivach R., Gluehmann M.,				
RA	Avila H., Bashan A., Bartels H., Auerbach T., Jacobl C., Hartsch T.,				
RA	Yonath A., Franceschi F.;				
RT	"Crystals structures of complexes of the small ribosomal subunit with				

```
DR PDB; 1JUG; 25-SEP-00.
DR PDB; 1QD7; 17-JAN-01.
DR PDB; 1HR0; 24-JAN-01.
DR PDB; 1HNW; 21-FEB-01.
DR PDB; 1HNW; 21-FEB-01.
DR PDB; 1HNZ; 21-FEB-01.
DR PDB; 1194; 12-APR-01.
DR PDB; 1195; 12-APR-01.
DR PDB; 1196; 12-APR-01.
DR PDB; 1197; 12-APR-01.
DR PDB; 1GIX; 01-JUN-01.
DR PDB; 1IBK; 04-MAY-01.
DR PDB; 1IBL; 04-MAY-01.
DR PDB; 1IBM; 04-MAY-01.
DR PDB; 1JGO; 14-SEP-01.
DR PDB; 1JGP; 14-SEP-01.
DR PDB; 1JGQ; 14-SEP-01.
DR PDB; 1LIU; 22-MAR-02.
DR PDB; 1JSE; 12-APR-02.
DR PDB; 1N32; 29-NOV-02.
DR HAMAP; MF 01306; -; 1.
DR InterPro; IPR001912; Ribosomal S4.
DR InterPro; IPR005709; Ribosomal_S4_b/o.
DR InterPro; IPR002942; S4.
DR Pfam; PF00163; Ribosomal_S4; 1.
DR Pfam; PF01479; S4; 1.
DR SMART; SMC0363; S4; 1.
DR TIGRFAMs; TIGR01017; rpsD_bact; 1.
DR PROSITE; PS00632; RIBOSOMAL_S4; 1.
DR PROSITE; PS00889; S4; 1.
KW Ribosomal protein; RNA-binding; rRNA-binding; Zinc; Metal-binding;
Zinc-finger; 3D-structure.
FT INIT MET 0 0
FT DOMAIN 98 160 S4 RNA-BINDING.
FT ZN FING 8 30 C4-TYPE.
FT METAL 8 8 ZINC (PROBABLE).
FT METAL 11 11 ZINC (PROBABLE).
FT METAL 25 25 ZINC (PROBABLE).
FT METAL 30 30 ZINC (PROBABLE).
FT CONFLICT 11 11 C -> S (IN REF. 2).
FT CONFLICT 25 25 C -> D (IN REF. 2).
SQ SEQUENCE 208 AA; 24193 MW; C75BA0458110E78 CRC64;

Query Match 30.6%; Score 75; DB 1; Length 208;
Best Local Similarity 34.1%; Pred. No. 0.036;
Matches 15; Conservative 9; Mismatches 20; Indels 0; Gaps 0;

QY 7 RADELVFLQGLAESRQAKRLIMAGKVTLTNNSTTIPRLKPG 50
Db 99 RLDNVVYRLGFAVSRQARQLVRGHITVNGRRVDLPYRVRPG 142

RESULT 9
RS4_STAAM STANDARD; PRT; 200 AA.
AC Q99TE4;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE 30S ribosomal protein S4.
GN RPSD OR SAV1719 OR SA1540.1 OR SAS052 OR MW1662.
OS Staphylococcus aureus (strain Mu50 / ATCC 700699),
OS Staphylococcus aureus (strain N315), and
OS Staphylococcus aureus (strain MW2).
OC Bacteria; Firmicutes; Bacillales; Staphylococcus.
OX NCBI_TaxID=158878, 158879, 196620;
[1]
RN SEQUENCE FROM N.A.
RP STRAIN=Mu50 / ATCC 700699, and N315;
RX MEDLINE=21311952; PubMed=11418146;
RA Kuroda M., Ohta T., Uchiyama I., Baba T., Yuzawa H., Kobayashi I.,
RA Cui L., Oguchi A., Aoki K.-I., Nagai Y., Iian J.-O., Ito T.,
RA Kanamori M., Matsumaru H., Maruyama A., Murakami H., Hoshyama A.,

Mizutani-Ui Y., Takahashi N.K., Sawano T., Inoue R.-I., Kaito C.,
Sekimizu K., Hirakawa H., Kuhara S., Goto S., Yabuzaki J.,
Kanehisa M., Yamashita A., Oshima K., Furuya K., Yoshino C., Shiba T.,
Hattori M., Ogasawara N., Hayashi H., Hiramatsu K.;
"Whole genome sequencing of methicillin-resistant Staphylococcus
aureus.";
RL Lancet 357:1225-1240(2001).
[2]
RN SEQUENCE FROM N.A.
RC STRAIN=MM2;
RX MEDLINE=22040717; PubMed=12044378;
RA Baba T., Takeuchi F., Kuroda M., Yuzawa H., Aoki K.-I., Oguchi A.,
RA Nagai Y., Iwama N., Asano K., Naimi T., Kuroda H., Cui L.,
RA Yamamoto K., Hiramatsu K.;
"Genome and virulence determinants of high virulence community-
acquired MRSA.";
RL Lancet 359:1819-1827(2002).
CC -!- FUNCTION: One of the primary rRNA binding proteins, it binds
CC directly to 16S rRNA where it nucleates assembly of the body of
CC the 30S subunit (By similarity).
CC -!- FUNCTION: With S5 and S12 plays an important role in translational
CC accuracy (By similarity).
CC -!- SUBUNIT: Part of the 30S ribosomal subunit. Contacts protein S5.
CC The interaction surface between S4 and S5 is involved in control
CC of translational fidelity (By similarity).
CC -!- SIMILARITY: Contains 1 S4 RNA-binding domain.
CC -!- SIMILARITY: Belongs to the S4P family of ribosomal proteins.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
CC EMBL; AP003363; BAB57881.1; -.
CC EMBL; AP003134; BAB42808.1; -.
CC EMBL; AP004827; BAB95527.1; -.
CC FIR; C89956; C89956.
CC HSPP; P81288; IC05.
CC HAMAP; MF 01306; -; 1.
CC InterPro; IPR001912; Ribosomal S4.
CC InterPro; IPR005709; Ribosomal_S4_b/o.
CC InterPro; IPR002942; S4.
CC Pfam; PF00163; Ribosomal_S4; 1.
CC Pfam; PF01479; S4; 1.
CC SMART; SMC0363; S4; 1.
CC TIGRFAMs; TIGR01017; rpsD_bact; 1.
CC PROSITE; PS00632; RIBOSOMAL_S4; FALSE_NEG.
CC PROSITE; PS00889; S4; 1.
KW Ribosomal protein; RNA-binding; rRNA-binding; Complete proteome.
FT DOMAIN 92 155 S4 RNA-BINDING.
SQ SEQUENCE 200 AA; 23013 MW; E4BBB131EF37A6ED CRC64;

Query Match 30.2%; Score 74; DB 1; Length 200;
Best Local Similarity 36.4%; Pred. No. 0.046;
Matches 16; Conservative 8; Mismatches 20; Indels 0; Gaps 0;

QY 7 RADELVFLQGLAESRQAKRLIMAGKVTLTNNSTTIPRLKPG 50
Db 93 RLDVAVYSLGATRTQARQLVNHGHLVDGKRDIPSYSVKPG 136

RESULT 10
RS4_BACAA STANDARD; PRT; 200 AA.
AC Q81KT2;
DT 15-MAR-2004 (Rel. 43, Created)
DT 15-MAR-2004 (Rel. 43, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE 30S ribosomal protein S4.
GN RPSD OR BA4908.
```

OS	Bacillus anthracis (strain Ames).
OC	Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX	NCBI_TaxID=198094;
RN	[1]
RP	SEQUENCE FROM N.A.
RX	MEDLINE=22608414; PubMed=1271629;
RA	Read T.D., Peterson S.N., Tourasse N., Baillie L.W., Paulsen I.T.,
RA	Nelson K.E., Tetzelin H., Fouts D.E., Eisen J.A., Gill S.R.,
RA	Kapral V., Bhattacharya A., Reznik G., Mikhailova N., Lapidus A.,
RA	Holzapfel E.K., Ostad O.A., Helgason E., Ristone J., Wu M.,
RA	Kolonay J.F., Beanan M.J., Dodson R.J., Brinkac L.M., Gwinn M.,
RA	DeBoy R.T., Madpu R., Doherty S.C., Durkin A.S., Haft D.H.,
RA	Nelson W.C., Peterson J.D., Pop M., Khouri H.M., Radune D.,
RA	Benton J.L., Mahamoud Y., Jiang L., Hance I.R., Weidman J.F.,
RA	Berry K.J., Plaut R.D., Wolf A.M., Watkins K.L., Nieman W.C.,
RA	Hazen A., Cline R., Redmond C., Twiss J.E., White O., Salzberg S.L.,
RA	Thomson B., Friedlander A.M., Koehler T.M., Hanna P.C., Kolsto A.-B.,
RA	Fraser C.M.;
RT	"The genome sequence of Bacillus anthracis Ames and comparison to
RT	closely related bacteria.";
RL	Nature 423:81-86(2003)
CC	-!- FUNCTION: One of the primary rRNA binding proteins, it binds
CC	directly to 16S rRNA where it nucleates assembly of the body of
CC	the 30S subunit (By similarity).
CC	-!- FUNCTION: With S5 and S12 plays an important role in translational
CC	accuracy (By similarity).
CC	-!- SUBUNIT: Part of the 30S ribosomal subunit. Contacts protein S5.
CC	The interaction surface between S4 and S5 is involved in control
CC	of translational fidelity (By similarity).
CC	-!- SIMILARITY: Belongs to the S4P family of ribosomal proteins.
CC	-!- SIMILARITY: Contains 1 S4 RNA-binding domain.
CC	-----
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC	the European Bioinformatics Institute. There are no restrictions on its
CC	use by non-profit institutions as long as its content is in no way
CC	modified and this statement is not removed. Usage by and for commercial
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC	or send an email to license@isb-sib.ch).
CC	-----
EMBL	AE017039; AAP28592.1; -
DR	TIGR; BA4908; -
DR	HAWAP; MF_01306; -; 1.
DR	InterPro; IPR001912; Ribosomal_S4.
DR	InterPro; IPR005709; Ribosomal_S4_b/o.
DR	InterPro; IPR002942; S4.
DR	Pfam; PF00163; Ribosomal_S4; 1.
DR	Pfam; PF01479; S4; 1.
DR	SMART; SM00363; S4; 1.
DR	TIGRFAMS; TIGR01017; rpsD_bact; 1.
DR	PROSITE; PS00632; RIBOSOMAL_S4; 1.
DR	PROSITE; PS00889; S4; 1.
KW	Ribosomal protein; rRNA-binding; Complete proteome.
FT	DOMAIN 92 152 S4 RNA-binding.
SQ	SEQUENCE 200 AA; 22996 MW; 51552DC60123FE6 CRC64;
	Query Match 29.8%; Score 73; DB 1; Length 200;
	Best Local Similarity 34.8%; Pred. No. 0.061;
	Matches 16; Conservative 8; Mismatches 22; Indels 0; Gaps 0;
QY	5 KVRADLVFLOGLAESREQAKRLIMAGKVTLTNNSTTIPLRLEKPG 50
	: : : : : : : : :
DB	91 EARLDNLVYRMGLARTTRAAQLVNHGHIHMDGARVDIPSYRVKPG 136
	: : : : : : : : :
RESULT 11	
ID	RS4_BACCR STANDARD; PRT; 200 AA.
AC	Q817A4;
DT	15-MAR-2004 (Rel. 43, Created)
DT	15-MAR-2004 (Rel. 43, Last sequence update)
DT	15-MAR-2004 (Rel. 43, Last annotation update)
DE	30S ribosomal protein S4.
GN	RPSD OR BC4555.

CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC -----
CC EMBL; AP005275; BAB97956.1; -.
CC HAMAP; MF_01306; -; 1.
CC InterPro; IPR001912; Ribosomal_S4.
CC InterPro; IPR005709; Ribosomal_S4_b/o.
CC InterPro; IPR002942; S4.
CC Pfam; PF00163; Ribosomal_S4; 1.
CC DR TIGRFAMS; TIGR01017; rpsd_bact; 1.
CC DR PROSITE; PS00632; RIBOSOMAL_S4; 1.
CC DR PROSITE; PS00889; S4; 1.
CC KW Ribosomal protein; rRNA-binding; Complete proteome.
FT DOMAIN 91 151
SQ SEQUENCE 201 AA; 23329 MW; 9D87D47E6F9A644D CRC64;

Query Match 28.6%; Score 70; DB 1; Length 201;
Best Local Similarity 35.1%; Pred. No. 0.15;
Matches 13; Conservative 9; Mismatches 15; Indels 0; Gaps 0;

QY 7 RADELVFLQGLAESRQAKRLIMAGKVTLTNNSTTIP 43

Db 92 RLDNVVYRAGLANTRQARQLVSHGHFTVNGKAIDVP 128

RESULT 15

RS4_THEVO STANDARD; PRT; 200 AA.
ID RS4_THEVO
AC Q97B95;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE 30S ribosomal protein S4P.
GN RPS4P OR TW0562 OR TWG0550709.
OS Thermoplasma volcanium.
OC Archaea; Euryarchaeota; Thermoplasmata; Thermoplasmatales;
OC Thermoplasmataceae; Thermoplasma.
OX NCBI_TaxID=50339;
RN [1]
RN SEQUENCE FROM N.A.
RC STRAIN=GSS1 / DSM 4299 / JCM 9571;
RX MEDLINE=20570466; PubMed=11121031;
RA Kawashima T., Amano N., Koike H., Makino S.-I., Higuchi S.,
RA Kawashima-Ohya Y., Watanabe K., Yamazaki M., Kanehori K., Kawamoto T.,
RA Nunoshiba T., Yamamoto Y., Aramaki H., Makino K., Suzuki M.,
RT "Archaeal adaptation to higher temperatures revealed by genomic
sequence of Thermoplasma volcanium."
CC Proc. Natl. Acad. Sci. U.S.A. 97:14257-14262(2000).
CC -!- FUNCTION: One of the primary rRNA binding proteins, it binds
CC directly to 16S rRNA where it nucleates assembly of the body of
CC the 30S subunit (By similarity).
CC -!- FUNCTION: With S5 and S12 plays an important role in translational
CC accuracy (By similarity).
CC -!- SUBUNIT: Part of the 30S ribosomal subunit. Contacts protein S5.
CC The interaction surface between S4 and S5 is involved in control
CC of translational fidelity (By similarity).
CC -!- SIMILARITY: Contains 1 S4 RNA-binding domain.
CC -!- SIMILARITY: Belongs to the S4P family of ribosomal proteins.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC EMBL; AP000992; BAB59704.1; -.
CC HAMAP; MF_01306; -; 1
CC InterPro; IPR001912; Ribosomal_S4.

DR InterPro; IPR005710; Ribosomal_S4/9.

DR InterPro; IPR002942; S4.

DR Pfam; PF00163; Ribosomal_S4; 1.

DR SMART; PF01479; S4; 1.

DR SMART; SM00363; S4; 1.

DR TIGRFAMS; TIGR01018; rpsd_arch; 1.

DR PROSITE; PS00632; RIBOSOMAL_S4; 1.

DR PROSITE; PS00889; S4; 1.

KW Ribosomal protein; rRNA-binding; Complete proteome.

FT DOMAIN 106 170 S4 RNA-BINDING.

SQ SEQUENCE 200 AA; 23161 MW; 71DBDE99D5F122F7 CRC64;

Query Match 28.2%; Score 69; DB 1; Length 200;

Best Local Similarity 37.8%; Pred. No. 0.19;

Matches 14; Conservative 8; Mismatches 15; Indels 0; Gaps 0;

QY 7 RADELVFLQGLAESRQAKRLIMAGKVTLTNNSTTIP 43

Db 107 RLQTIIVFKKNLALSEKQARQLITHGHITVNGRRVIVP 143

RESULT 16

RR4_TAKLE STANDARD; PRT; 202 AA.

ID RR4_TAKLE

AC P59150;

DT 28-FEB-2003 (Rel. 41, Created)

DT 28-FEB-2003 (Rel. 41, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE Chloroplast 30S ribosomal protein S4.

GN RPS4.

OS Takakia lepidiozioides (Moss).

OG Chloroplast.

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Bryophyta;

OC Takakiopsida; Takakiales; Takakiaceae; Takakia.

OX NCBI_TaxID=37425;

RN [1]

RN SEQUENCE FROM N.A.

RC TISSUE=Gametophyte;

RA Capesius I., Bloecher R.;

RT "A molecular approach to bryophyte systematics.";

RT Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.

CC -!- FUNCTION: One of the primary rRNA binding proteins, it binds

CC directly to 16S rRNA where it nucleates assembly of the body of

CC the 30S subunit (By similarity).

CC -!- FUNCTION: With S5 and S12 plays an important role in translational

CC accuracy (By similarity).

CC -!- SUBUNIT: Part of the 30S ribosomal subunit. Contacts protein S5.

CC The interaction surface between S4 and S5 is involved in control

CC of translational fidelity (By similarity).

CC -!- SUBCELLULAR LOCATION: Chloroplast.

CC -!- SIMILARITY: Contains 1 S4 RNA-binding domain.

CC -!- SIMILARITY: Belongs to the S4P family of ribosomal proteins.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

CC -----

CC EMBL; AJ269687; CAC80628.1; -.

CC HAMAP; MF_01306; -; 1.

CC InterPro; IPR001912; Ribosomal_S4.

CC InterPro; IPR005709; Ribosomal_S4_b/o.

CC InterPro; IPR002942; S4.

CC Pfam; PF00163; Ribosomal_S4; 1.

CC Pfam; PF01479; S4; 1.

CC SMART; SM00363; S4; 1.

CC TIGRFAMS; TIGR01017; rpsd_bact; 1.

CC PROSITE; PS00632; RIBOSOMAL_S4; 1.

CC PROSITE; PS00889; S4; 1.

CC Ribosomal protein; rRNA-binding; Chloroplast.

KW

```
FT DOMAIN 90 153 S4 RNA-BINDING.
SQ SEQUENCE 202 AA; 23395 MW; FD942561BE324003 CRC64;

Query Match 28.2%; Score 69; DB 1; Length 202;
Best Local Similarity 30.4%; Pred. No. 0.2;
Matches 14; Conservative 12; Mismatches 20; Indels 0; Gaps 0;

QY 5 KVRDELVFLQGLASRQAKRLIMAGKVTLTNNSTTPIRLKPKG 50
DB 89 EMRLDNTIFRLGMAPTIPCARQLVNHGHISMNNIIDIPSYNREP 134

RESULT 17
ID KIN2 YEAST STANDARD; PRT; 1147 AA.
AC P13186; Q12384;
DT 01-JAN-1990 (Rel. 13, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Protein kinase KIN2 (EC 2.7.1.37).
GN KIN2 OR YLR096W OR L8004.3.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4332;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=87317589; PubMed=2957690;
RA Levin D.E., Hammond C.I., Ralston R.O., Bishop J.M.;
RT "Two yeast genes that encode unusual protein kinases.";
RL Proc. Natl. Acad. Sci. U.S.A. 84:6035-6039(1987).
RN [2]
RP SEQUENCE FROM N.A.
RX STRAIN=S288C / AB972;
RA Johnston M., Hillier L., Riles L., Albermann K., Andre B., Ansoerge W.,
RA Benes V., Brueckner M., Delius H., Dubois E., Duesterhoeft A.,
RA Etian K.-D., Floeth M., Goffeau A., Hebling U., Heumann K.,
RA Heuss-Neitzel D., Hilbert H., Hilger F., Kleine K., Koetter P.,
RA Louis E.J., Messenguy F., Mewes H.-W., Miosga T., Mostl D.,
RA Mueller-Auer S., Nentwich U., Obermaier B., Piravandi E., Pohl T.M.,
RA Portetelle D., Fumelle B., Rechmaier S., Rieger M., Rinke M., Rose M.,
RA Scharte M., Scherens B., Scholler P., Schwager C., Schwarz S.,
RA Underwood A.P., Urrestazu L.A., Vandenbol M., Verhasselt P.,
RA Vierendeels F., Voet M., Volckaert G., Voss H., Wambutt R., Wedler E.,
RA Wedler H., Zimmermann F.K., Zollner A., Hani J., Hobeisel J.D.;
RA "The nucleotide sequence of Saccharomyces cerevisiae chromosome XII.";
RT Nature 387:87-90(1997).
CC -!- FUNCTION: This protein is probably a serine/threonine protein
CC Kinase.
CC -!- CATALYTIC ACTIVITY: ATP + a protein = ADP + a phosphoprotein.
CC -!- SIMILARITY: Belongs to the Ser/Thr family of protein kinases. NIM1
CC subfamily.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
CC EMBL; M69018; AAA34723.1; -
CC EMBL; Z73268; CAA97659.1; -
CC EMBL; Z73269; CAA97661.1; -
CC EMBL; U53876; AAB67540.1; -
CC PIR; S64930; S64930.
CC HSSP; Q63450; IA06.
CC GermOnline; 142158; -.
CC SGD; S0004086; KIN2.
CC GO; GO:0005624; C:membrane fraction; IDA.
CC GO; GO:0004672; F:protein kinase activity; IDA.
CC InterPro; IPR001772; Kinase_Cterm.
```

```
DR InterPro; IPR000719; Prot Kinase.
DR InterPro; IPR008271; Ser Thr_pkin_AS.
DR InterPro; IPR002290; Ser_thr_pkinase.
DR InterPro; IPR001345; Tyr_pkinase.
DR Pfam; PF02149; KAI; 1.
DR Pfam; PF00069; pkinase; 1.
DR PRINTS; PD000001; Prot Kinase.
DR ProDom; PD000001; Prot Kinase; 1.
DR SMART; SM00220; S_TKC; 1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
DR PROSITE; PS00111; PROTEIN_KINASE_DOM; 1.
KW Transferase; Serine/threonine-protein kinase; ATP-binding.
FT DOMAIN 99 377 PROTEIN KINASE.
FT NP_BIND 105 113 ATP (BY SIMILARITY).
FT BINDING 128 128 ATP (BY SIMILARITY).
FT ACT_SITE 248 248 BY SIMILARITY.
FT DOMAIN 527 536 POLY-GLN.
FT CONFLICT 216 217 OH -> HD (IN REF. 1).
FT CONFLICT 675 707 OEPLPERPPYMSKNEISKVPKSHRTISD ->
FT SGTYSSENLEQHICQNMKFKYRKAIVLYQT (IN
FT REF. 1).
FT CONFLICT 756 758 NAE -> KRQ (IN REF. 1).
FT CONFLICT 805 805 P -> PLSVP (IN REF. 1).
FT CONFLICT 1034 1037 ATNT -> TTNSI (IN REF. 1).
FT CONFLICT 1041 1042 NS -> KT (IN REF. 1).
FT SEQUENCE 1147 AA; 128338 MW; AC5660BF3CA69600 CRC64;

Query Match 28.2%; Score 69; DB 1; Length 1147;
Best Local Similarity 31.1%; Pred. No. 1.1;
Matches 14; Conservative 12; Mismatches 19; Indels 0; Gaps 0;

QY 5 KVRDELVFLQGLASRQAKRLIMAGKVTLTNNSTTPIRLKPK 49
DB 515 KLRORQALALQAQAOQOQQOQQVALGKVALNNSPDIMTKRSP 559

RESULT 18
RS4_MYCPU
ID_RS4_MYCPU STANDARD; PRT; 199 AA.
AC Q98PKG;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE 30S ribosomal protein S4.
GN RPSD OR MIPU_7160.
OS Mycoplasma pulmonis.
OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.
OC NCBI_TaxID=2107;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=UAB CTIP;
RX MEDLINE=21267165; PubMed=11353084;
RA Chambaud I., Heilig R., Ferris S., Barbe V., Samson D., Galisson F.,
RA Moszer I., Dybvig K., Wroblewski H., Viari A., Rocha E.P.C.,
RA Blanchard A.;
RT "The complete genome sequence of the murine respiratory pathogen
RT Mycoplasma pulmonis.";
RL Nucleic Acids Res. 29:2145-2153(2001).
CC -!- FUNCTION: One of the primary rRNA binding proteins, it binds
CC directly to 16S rRNA where it nucleates assembly of the body of
CC the 30S subunit (By similarity).
CC -!- FUNCTION: With S5 and S12 plays an important role in translational
CC accuracy (By similarity).
CC -!- SUBUNIT: Part of the 30S ribosomal subunit. Contacts protein S5.
CC The interaction surface between S4 and S5 is involved in control
CC of translational fidelity (By similarity).
CC -!- SIMILARITY: Contains 1 S4 RNA-binding domain.
CC -!- SIMILARITY: Belongs to the S4p family of ribosomal proteins.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
```



```

TIGR; MG311; -.
DR HAWAP; MF_01306; -; 1.
DR InterPro; IPR001912; Ribosomal_S4.
DR InterPro; IPR005709; Ribosomal_S4_b/o.
DR InterPro; IPR002942; S4.
DR Pfam; PF00163; Ribosomal_S4; 1.
DR Pfam; PF01479; S4; 1.
DR SMART; SM00363; S4; 1.
DR TIGRFAMs; TIGR01017; rpsD_bact; 1.
DR PROSITE; PS00632; RIBOSOMAL_S4; 1.
DR PROSITE; PS50889; S4; 1.
KW Ribosomal protein; RNA-binding; rRNA-binding; Complete proteome.
FT DOMAIN 95 158 S4 RNA-BINDING.
SQ SEQUENCE 205 AA; 23947 MW; 3F2861E791339118 CRC64;

Query Match      27.3%; Score 67; DB 1; Length 205;
Best Local Similarity 29.5%; Pred. No. 0.35;
Matches 13; Conservative 11; Mismatches 20; Indels 0; Gaps 0

QY 7 RADELVFLQGLAESREQXRLIMAGKVTLTNNSTIPLRLKPG 50
   |||::||::||::||::||::||::||::||::||::||
DB 96 RLDNIVYRMGFAPTRKSARQWNHGVILNDQTVDTSIIINPG 139

RESULT 22
RS4_METJUA          STANDARD;              PRT;    187 AA.
ID   RS4_METJUA
AC   P54020;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   30S ribosomal protein S4P.
GN   RPS4P OR MJ0190.
OS   Methanococcus jannaschii.
OC   Archaea; Euryarchaeota; Methanococci; Methanococcales;
OX   NCBI_TaxId=2190;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=JAL-1 / DSM 2661 / ATCC 43067;
RX   MEDLINE=96337999; PubMed=868808;
RA   Buit C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
RA   Sutton G.G., Blake J.A., FitzGerald L.M., Clayton R.A., Gocayne J.D.,
RA   Kervilange A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.I.,
RA   Overbeek R., Kirkness E.P., Weinstock K.G., Merrick J.M., Glodek A.,
RA   Scott J.L., Geoghegan N.S.M., Weidman J.F., Fuhrmann J.B., Nguyen D.,
RA   Uterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
RA   Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,
RA   Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
RT   "Complete genome sequence of the methanogenic archaeon, Methanococcus
RL   jannaschii.";
RL   Science 273:1058-1073(1996).
CC   -!- FUNCTION: One of the primary rRNA binding proteins, it binds
CC   directly to 16S rRNA where it nucleates assembly of the body of
CC   the 30S subunit [By similarity].
CC   -!- FUNCTION: With S5 and S12 plays an important role in translational
CC   accuracy [By similarity].
CC   -!- SUBUNIT: Part of the 30S ribosomal subunit. Contacts protein S5.
CC   The interaction surface between S4 and S5 is involved in control
CC   of translational fidelity [By similarity].
CC   -!- SIMILARITY: Contains 1 S4 RNA-binding domain.
CC   -!- SIMILARITY: Belongs to the S4P family of ribosomal proteins.
CC   -----
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation at
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial/
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -----
EMBL; U67475; AAB98170.1; -.
DR FIC; G64323; G64323.

```

```

DR HSSP; P81288; LC05.
DR TIGR; M0190; -.
DR HAMAP; MF_01306; -, 1.
DR InterPro; IPR001912; Ribosomal_S4.
DR InterPro; IPR005710; Ribosomal_S4/9.
DR InterPro; IPR002942; S4.
DR Pfam; PF01479; S4; 1.
DR SMART; SM00363; S4; 1.
DR TIGRFAMs; TIGR01018; rpsd arch; 1.
DR PROSITE; PS00632; RIBOSOMAL_S4; 1.
DR PROSITE; PS0889; S4; 1.
KW Ribosomal protein; rRNA-binding; Complete proteome.
FT DOMAIN 105 174 S4 RNA-BINDING.
SQ SEQUENCE 187 AA; 22050 MW; 7D4C51389D5C8A12 CRC64;

Query Match 26.9%; Score 66; DB 1; Length 187;
Best Local Similarity 37.8%; Pred. No. 0.43; Mismatches 0; Gaps 0;
Matches 14; Conservative 7;

QY 7 RADELVFLQGLAESREQAKRLIMAGKVTLTNNSTTIP 43
DB 106 RLQTLVFKGLARTPRQARQLIVGHIANGRVVTAP 142

RESULT 23
RS4 LISIN STANDARD; PRT; 200 AA.
AC Q92B22;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE 30S ribosomal protein S4.
GN RPSD OR LIN1638.
OS Listeria innocua.
OC Bacteria; Firmicutes; Bacillales; Listeriaceae; Listeria.
OX NCBI_TaxID=1642;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CLIP 11262 / Serovar 6a;
RX MEDLINE=21537279; PubMed=11679669;
RA Glaser P., Frangeul L., Buchrieser C., Rusniok C., Amend A.,
RA Baquero F., Berche P., Bloecher H., Brandt P., Chakraborty T.,
RA Charbit A., Chetouani F., Couve E., de Daruvar A., Dehoux P.,
RA Domann E., Dominguez-Bernal G., Duchaud E., Durant L., Dussurget O.,
RA Etian K.-D., Fsihi H., Garcia-del Portillo F., Garrido P.,
RA Gautier L., Goebel W., Gomez-Lopez N., Hain T., Hauf J., Jackson D.,
RA Jones L.-M., Kaerst U., Kreft J., Kuhn M., Kunst F., Kurapkat G.,
RA Madueno E., Maltournam A., Mata Vicente J., Ng E., Nedjari H.,
RA Nordsiek G., Novella S., de Pablos B., Perez-Diaz J.-C., Purcell R.,
RA Remmel B., Rose M., Schluster T., Simoes N., Tierrez A.,
RA Vazquez-Boland J.-A., Voss H., Wehland J., Cossart P.;
RT "Comparative genomics of Listeria species.";
RL Science 294:849-852(2001).
CC -!- FUNCTION: One of the primary rRNA binding proteins, it binds
CC directly to 16S rRNA where it nucleates assembly of the body of
CC the 30S subunit (By similarity).
CC -!- FUNCTION: With S5 and S12 plays an important role in translational
CC accuracy (By similarity).
CC -!- SUBUNIT: Part of the 30S ribosomal subunit. Contacts protein S5.
CC The interaction surface between S4 and S5 is involved in control
CC of translational fidelity (By similarity).
CC -!- SIMILARITY: Contains 1 S4 RNA-binding domain.
CC -!- SIMILARITY: Belongs to the S4P family of ribosomal proteins.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AL596169; CAC96869.1; -.

```

```

DR PIR; AE1637; AE1637.
DR ListList; LIN01638; -.
DR HAMAP; MF_01306; -, 1.
DR InterPro; IPR001912; Ribosomal_S4.
DR InterPro; IPR005709; Ribosomal_S4_b/o.
DR InterPro; IPR002942; S4.
DR Pfam; PF00163; Ribosomal_S4; 1.
DR Pfam; PF01479; S4; 1.
DR SMART; SM00363; S4; 1.
DR TIGRFAMs; TIGR01017; rpsd bact; 1.
DR PROSITE; PS00632; RIBOSOMAL_S4; 1.
DR PROSITE; PS0889; S4; 1.
KW Ribosomal protein; rRNA-binding; Complete proteome.
FT DOMAIN 92 170 S4 RNA-BINDING.
SQ SEQUENCE 200 AA; 22680 MW; 5114B9DB68970820 CRC64;

Query Match 26.9%; Score 66; DB 1; Length 200;
Best Local Similarity 35.1%; Pred. No. 0.46; Mismatches 8; Gaps 0;
Matches 13; Conservative 8;

QY 7 RADELVFLQGLAESREQAKRLIMAGKVTLTNNSTTIP 43
DB 93 RLIDNIVYRLGLARTPRARQLVNHGHITVDGKRVDP 129

RESULT 24
RS4 LISMO STANDARD; PRT; 200 AA.
AC Q8Y6T6;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE 30S ribosomal protein S4.
GN RPSD OR IMO1596.
OS Listeria monocytogenes.
OC Bacteria; Firmicutes; Bacillales; Listeriaceae; Listeria.
OX NCBI_TaxID=1639;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=EGD-e / Serovar 1/2a;
RX MEDLINE=21537279; PubMed=11679669;
RA Glaser P., Frangeul L., Buchrieser C., Rusniok C., Amend A.,
RA Baquero F., Berche P., Bloecher H., Brandt P., Chakraborty T.,
RA Charbit A., Chetouani F., Couve E., de Daruvar A., Dehoux P.,
RA Domann E., Dominguez-Bernal G., Duchaud E., Durant L., Dussurget O.,
RA Etian K.-D., Fsihi H., Garcia-del Portillo F., Garrido P.,
RA Gautier L., Goebel W., Gomez-Lopez N., Hain T., Hauf J., Jackson D.,
RA Jones L.-M., Kaerst U., Kreft J., Kuhn M., Kunst F., Kurapkat G.,
RA Madueno E., Maltournam A., Mata Vicente J., Ng E., Nedjari H.,
RA Nordsiek G., Novella S., de Pablos B., Perez-Diaz J.-C., Purcell R.,
RA Remmel B., Rose M., Schluster T., Simoes N., Tierrez A.,
RA Vazquez-Boland J.-A., Voss H., Wehland J., Cossart P.;
RT "Comparative genomics of Listeria species.";
RL Science 294:849-852(2001).
CC -!- FUNCTION: One of the primary rRNA binding proteins, it binds
CC directly to 16S rRNA where it nucleates assembly of the body of
CC the 30S subunit (By similarity).
CC -!- FUNCTION: With S5 and S12 plays an important role in translational
CC accuracy (By similarity).
CC -!- SUBUNIT: Part of the 30S ribosomal subunit. Contacts protein S5.
CC The interaction surface between S4 and S5 is involved in control
CC of translational fidelity (By similarity).
CC -!- SIMILARITY: Contains 1 S4 RNA-binding domain.
CC -!- SIMILARITY: Belongs to the S4P family of ribosomal proteins.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AL596169; CAC96869.1; -.

```

```
CC use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announcement/ or send an email to license@isb-sib.ch)
CC -----
CC EMBL; U30821; AAA81243.1; -
CC DR PIR; T06900; T06900.
CC DR HSSP; P81288; IC05.
CC DR HAMAP; MF_01306; -, 1.
CC DR InterPro; IPR001912; Ribosomal_S4.
CC DR InterPro; IPR005709; Ribosomal_S4_b/o.
CC DR InterPro; IPR002942; S4.
CC DR Pfam; PF00163; Ribosomal_S4; 1.
CC DR Pfam; PF01479; S4; 1.
CC DR SMART; SM00363; S4; 1.
CC DR TIGRFAMs; TIGR01017; rpsD_bact; 1.
CC DR PROSITE; PS00632; RIBOSOMAL_S4; 1.
CC DR PROSITE; PS00889; S4; 1.
CC KW Ribosomal protein; RNA-binding; rRNA-binding; Cyanelle.
CC FT DOMAIN 92 170 S4 RNA-BINDING.
CC SQ SEQUENCE 200 AA; 22735 MW; 511569D8689F9CC CRC64;

Query Match      26.9%; Score 66; DB 1; Length 200;
Best Local Similarity 35.1%; Pred.No. 0.46;
Matches 13; Conservative 8; Mismatches 16; Indels 0; Gaps 0;

QY 7 RADELVFLQLAESREQAKELIMAGKVTLTNNSTTP 43
    |||::|||::|||::|||::|||::|||::|||
Db 93 RLNDIVVLGLARTRAARQLVNHGHTVDGKRDVP 129

RESULT 25
RR4 CYAPA STANDARD; PRT; 201 AA.
AC P48133;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Cyanelle 30S ribosomal protein S4.
GN RPS4.
OS Cyanophora paradoxa.
OC Bacteria; Glaucocystophyceae; Cyanophoraceae; Cyanophora.
OG Cyanelle.
OX NCBI_TaxID=2762;
[1] _SEQUENCE FROM N.A.
RP STRAIN=UTEX LB 555 / Pringsheim;
RC Stirewalt V.L., Michalowski C.B., Loeffelhardt W., Bohnert H.J., Bryant D.A.;
RL "Nucleotide sequence of the cyanelle DNA from Cyanophora paradoxa.";
RT Plant Mol. Biol. Rep. 13:327-332(1995).
RN [2]
RP _SEQUENCE FROM N.A.
RC STRAIN=UTEX LB 555 / Pringsheim;
RC Loeffelhardt W., Stirewalt V.L., Michalowski C.B., Annarella M., Farley J.Y., Schlucher W.M., Chung S., Neumann-Spallart C., Steiner J.M., Jakowitsch J., Bohnert H.J., Bryant D.A.;
RT "The complete sequence of the cyanelle genome of Cyanophora paradoxa: the genetic complexity of a primitive plastid."
RL (in) Schenk H.E.A., Herrmann R., Jeon K.W., Mueller N.E., Schwemmler W. (eds.);
RL Eukaryotism and Symbiosis, pp.40-48, Springer-Verlag, Heidelberg (1997).
RL
CC -!- FUNCTION: One of the primary rRNA binding proteins, it binds directly to 16S rRNA where it nucleates assembly of the body of the 30S subunit (By similarity).
CC CC
CC -!- FUNCTION: With S5 and S12 plays an important role in translational accuracy (By similarity).
CC CC
CC -!- SUBUNIT: Part of the 30S ribosomal subunit. Contacts protein S5. The interaction surface between S4 and S5 is involved in control of translational fidelity (By similarity).
CC CC
CC -!- SURCELLULAR LOCATION: Cyanelle.
CC CC
CC -!- SIMILARITY: Contains 1 S4 RNA-binding domain.
CC CC
CC -!- SIMILARITY: Belongs to the S4P family of ribosomal proteins.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announcement/ or send an email to license@isb-sib.ch)
CC -----
CC EMBL; AE002148; AAP30910.1; -.
CC DR HSSP; P81288; IC05.
```

```
DR HAMAP; MF_01306; -; 1.
DR InterPro; IPR001912; Ribosomal_S4.
DR InterPro; IPR005709; Ribosomal_S4_b/o.
DR InterPro; IPR002942; S4.
DR Pfam; PF00163; Ribosomal_S4; 1.
DR Pfam; PF01479; S4; 1.
DR SMART; SM00363; S4; 1.
DR TIGRFAMs; TIGR01017; rpsD bact; 1.
DR PROSITE; PS00632; RIBOSOMAL_S4; 1.
DR PROSITE; PS00889; S4; 1.
KW Ribosomal protein; rRNA-binding; rRNA-binding; Zinc; Metal-binding;
FT DOMAIN 94 157 S4 RNA-BINDING.
SQ SEQUENCE 202 AA; 23124 MW; P929E0AF52848C2E CRC64;

Query Match 26.9%; Score 66; DB 1; Length 202;
Best Local Similarity 34.1%; Pred. No. 0.46; 22; Indels 0; Gaps 0;
Matches 15; Conservative 7; Mismatches 0;

OY 7 RADELVFLQGLAESRQAKRLIMAGKVTLTNNSTTPIRLLEKP 50
DB 95 RLDSLYVRAGFAETRAARQLVNHSHVNVNKKVIPSALVEVG 138

RESULT 27
RS4 THEME STANDARD; PRT; 209 AA.
AC QXLI13;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE 30S ribosomal protein S4.
GN RPSD OR TM1473
OS Thermotoga maritima.
OC Bacteria; Thermotogae; Thermotogales; Thermotogaceae; Thermotoga.
OX NCBI_TaxID=2336;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MSB8 / DSM 3109 / ATCC 43589;
RX MEDLINE=9287316; PubMed=10360571;
RA Nelson K.E., Clayton R.A., Gill S.R., Gwinn M.L., Dodson R.J.,
RA Haft D.H., Hickey E.K., Peterson J.D., Nelson W.C., Ketchum K.A.,
RA McDonald L., Utterback T.R., Malek J.A., Linher K.D., Garrett M.M.,
RA Stewart A.M., Cotton M.D., Pratt M.S., Phillips C.A., Richardson D.,
RA Heidelberg J., Sutton G.G., Fleischmann R.D., Eisen J.A., White O.,
RA Salzberg S.L., Smith H.O., Venter J.C., Fraser C.M.;
RT "Evidence for lateral gene transfer between Archaea and Bacteria from
RT genome sequence of Thermotoga maritima.";
RL Nature 399:323-329(1999).
CC -!- FUNCTION: One of the primary rRNA binding proteins, it binds
CC directly to 16S rRNA where it nucleates assembly of the body of
CC the 30S subunit (By similarity).
CC -!- FUNCTION: With S5 and S12 plays an important role in translational
CC accuracy (By similarity).
CC -!- COFACTOR: Zinc (By similarity).
CC -!- SUBUNIT: Part of the 30S ribosomal subunit. Contacts protein S5.
CC The interaction surface between S4 and S5 is involved in control
CC of translational fidelity (By similarity).
CC -!- SIMILARITY: Contains 1 S4 RNA-binding domain.
CC -!- SIMILARITY: Belongs to the S4P family of ribosomal proteins.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AE001798; AAD36541.1; -.
CC DR PIR; B72247; B72247.
CC DR HSP; P81288; 1C05.
CC TIGR; TM1473; -.
CC HAMAP; MF_01306; -; 1.

DR HAMAP; MF_01306; Ribosomal_S4.
DR InterPro; IPR005709; Ribosomal_S4_b/o.
DR InterPro; IPR002942; S4.
DR Pfam; PF00163; Ribosomal_S4; 1.
DR Pfam; PF01479; S4; 1.
DR SMART; SM00363; S4; 1.
DR TIGRFAMs; TIGR01017; rpsD bact; 1.
DR PROSITE; PS00632; RIBOSOMAL_S4; 1.
DR PROSITE; PS00889; S4; 1.
KW Ribosomal protein; rRNA-binding; rRNA-binding; Zinc; Metal-binding;
FT DOMAIN 98 161 S4 RNA-BINDING.
SQ SEQUENCE 209 AA; 24720 MW; 441E442403A56721 CRC64;

Query Match 26.9%; Score 66; DB 1; Length 209;
Best Local Similarity 31.1%; Pred. No. 0.48; 22; Indels 0; Gaps 0;
Matches 14; Conservative 9; Mismatches 9;

OY 5 KVRADLVFLQGLAESRQAKRLIMAGKVTLTNNSTTPIRLLEKP 49
DB 97 EARLDNVVVRMGFAINRQARQLVNHGHFLVNGCKVNPISYLLRP 141

RESULT 28
RR4 MESVI STANDARD; PRT; 205 AA.
ID -RR4 MESVI
AC QSMUNO;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Chloroplast 30S ribosomal protein S4.
GN RPS4.
OS Mesostigma viride.
OC Chloroplast.
OC Eukaryota; Viridiplantae; Streptophyta; Mesostigmatophyceae;
OC Mesostigmatales; Mesostigmataceae; Mesostigma.
OX NCBI_TaxID=41882;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NIES-296;
RX MEDLINE=20150907; PubMed=10688199;
RA Leneux C., Otis C., Turmel M.;
RT "Ancestral chloroplast genome in Mesostigma viride reveals an early
RT branch of green plant evolution.";
RL Nature 403:649-652(2000).
CC -!- FUNCTION: One of the primary rRNA binding proteins, it binds
CC directly to 16S rRNA where it nucleates assembly of the body of
CC the 30S subunit (By similarity).
CC -!- FUNCTION: With S5 and S12 plays an important role in translational
CC accuracy (By similarity).
CC -!- SUBUNIT: Part of the 30S ribosomal subunit. Contacts protein S5.
CC The interaction surface between S4 and S5 is involved in control
CC of translational fidelity (By similarity).
CC -!- SUBCELLULAR LOCATION: Chloroplast.
CC -!- SIMILARITY: Contains 1 S4 RNA-binding domain.
CC -!- SIMILARITY: Belongs to the S4P family of ribosomal proteins.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AF166114; AAF43870.1; -.
CC DR HSP; P81288; 1C05.
CC HAMAP; MF_01306; -; 1.
```

```

DR InterPro; IPR001912; Ribosomal_S4.
DR InterPro; IPR005709; Ribosomal_S4_b/o.
DR Pfam; PF00163; Ribosomal_S4.
DR Pfam; PF01479; S4; 1.
DR SMART; SM00363; S4; 1.
DR TIGRFAMs; TIGR01017; rpsD_bact; 1.
DR PROSITE; PS00632; RIBOSOMAL_S4; 1.
DR PROSITE; PS00889; S4; 1.
KW Ribosomal protein; RNA-binding; rRNA-binding; Chloroplast.
FT DOMAIN 93 156 S4 RNA-BINDING.
SQ SEQUENCE 205 AA; 23365 MW; E3C823ACBC3B2812 CRC64;

Query Match 26.5%; Score 65; DB 1; Length 205;
Best Local Similarity 35.6%; Pred. No. 0.63;
Matches 16; Conservative 8; Mismatches 21; Indels 0; Gaps 0;

QY 5 KVADELVFLQGLAESRQAKRLIMAGKVTLTNNSTTIPRLKPG 49
Db 92 EMRLDNTVFRGLAPTATAGARQLVNHGHIMVNNRIVTIPSYKCKP 136

RESULT 29
RS4_STRPN ID - RS4 STRPN STANDARD; PRT; 203 AA.
AC Q97T69;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE 30S ribosomal protein S4.
GN RPSD OR SP0095 OR SPR0078.
OS Streptococcus pneumoniae, and
OS Streptococcus pneumoniae (strain ATCC BAA-255 / R6).
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1313, 171101;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC BAA-334 / TIGR4;
RX MEDLINE=21357209; PubMed=11463916;
RA Tettelin H., Nelson K.E., Paulsen I.T., Eisen J.A., Read T.D.,
RA Peterson S., Heidelberg J., DeBoy R.T., Haft D.H., Dodson R.J.,
RA Durkin A.S., White M., Kolonay J.F., Nelson W.C., Peterson J.D.,
RA Unayian L.A., Gwinn M., Salzberg S.L., Lewis M.R., Radune D.,
RA Holtzapple E., Khouri H., Wolf A.M., Uterback T.R., Hansen C.L.,
RA McDonald L.A., Feldblyum T.V., Angiolini S., Dickinson T., Hickey E.K.,
RA Holt I.E., Loftus B.J., Yang F., Smith H.O., Venter J.C.,
RA Dougherty B.A., Morrison D.A., Hollinghead S.K., Fraser C.M.;
RT "Complete genome sequence of a virulent isolate of Streptococcus
RT pneumoniae."
RL Science 293:498-506(2001).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC BAA-255 / R6;
RX MEDLINE=21429245; PubMed=11544234;
RA Hoskins J., Alborn W.E. Jr., Arnold J., Blaszcak L.C., Burgett S.,
RA DeHoff B.S., Estrem S.T., Fritz L., Fu D.-J., Fuller W., Geringer C.,
RA Gilmour R., Glass J.S., Khoja H., Kraft A.R., Lagace R.E.,
RA LeBlanc D.J., Lee L.N., Lefkowitz E.J., Lu J., Matsushima P.,
RA McArris S.M., McHenry M., McLeaster K., Mundy C.W., Nicas T.I.,
RA Norris F.H., O'Gara M., Peery R.B., Robertson G.T., Rockey P.,
RA Sun P.-M., Winkler M.E., Yang Y., Young-Bellido M., Zhao G.,
RA Zook C.A., Baltz R.H., Jaskunas S.R., Rostock P.R. Jr., Skatrud P.L.,
RA Glass J.I.;
RT "Genome of the bacterium Streptococcus pneumoniae strain R6."
RL J. Bacteriol. 183:5709-5717(2001).
CC -!- FUNCTION: One of the primary rRNA binding proteins, it binds
CC directly to 16S rRNA where it nucleates assembly of the body of
CC the 30S subunit (By similarity).
CC -!- FUNCTION: With S5 and S12 plays an important role in translational
CC accuracy (By similarity).
CC -!- SUBUNIT: Part of the 30S ribosomal subunit. Contacts protein S5.
CC The interaction surface between S4 and S5 is involved in control

```

```

CC of translational fidelity (By similarity).
CC -!- SIMILARITY: Contains 1 S4 RNA-binding domain.
CC -!- SIMILARITY: Belongs to the S4P family of ribosomal proteins.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
DR EMBL; AE007326; AAK74273.1; -.
DR EMBL; AE008392; AAK98882.1; -.
DR PIR; F97881; F97881.
DR PIR; H95009; H95009.
DR TIGR; SP0085; -.
DR HAMAP; MF_01306; -.
DR InterPro; IPR001912; Ribosomal_S4.
DR InterPro; IPR005709; Ribosomal_S4_b/o.
DR InterPro; IPR002942; S4.
DR Pfam; PF00163; Ribosomal_S4; 1.
DR Pfam; PF01479; S4; 1.
DR TIGRFAMs; TIGR01017; rpsD_bact; 1.
DR PROSITE; PS00632; RIBOSOMAL_S4; 1.
DR PROSITE; PS00889; S4; 1.
KW Ribosomal protein; RNA-binding; rRNA-binding; Complete proteome.
FT DOMAIN 93 156 S4 RNA-BINDING.
SQ SEQUENCE 203 AA; 23028 MW; A35CBE67F49FC6BD CRC64;

Query Match 26.1%; Score 64; DB 1; Length 203;
Best Local Similarity 31.8%; Pred. No. 0.83;
Matches 14; Conservative 8; Mismatches 22; Indels 0; Gaps 0;

QY 7 RADELVFLQGLAESRQAKRLIMAGKVTLTNNSTTIPRLKPG 50
Db 94 RLDNVVRLGLATTRQARQVFNHGHILVDGKRVDFSYRVTPG 137

RESULT 30
RS4_STRPN ID - RS4 STRPN STANDARD; PRT; 203 AA.
AC Q99XJ4;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE 30S ribosomal protein S4.
GN RPSD OR SPY2178 OR SPYM3_1833 OR SPS1829 OR SPYM18_2215.
OS Streptococcus pyogenes,
OS Streptococcus pyogenes (serotype M3), and
OS Streptococcus pyogenes (serotype M18).
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1314, 198466, 186103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SF370 / ATCC 700294 / Serotype M1;
RX MEDLINE=21192684; PubMed=1196296;
RA Ferretti J.J., McShan W.M., Ajdic D.J., Savic D.J., Savic G., Lyon K.,
RA Primeaux C., Sezate S., Suvarov A.N., Kenton S., Lai H.S., Lin S.P.,
RA Qian Y., Jia H.G., Najjar F.Z., Ren Q., Zhu H., Song L., White J.,
RA Yuan X., Clifton S.W., Roe B.A., McLaughlin R.;
RT "Complete genome sequence of an M1 strain of Streptococcus pyogenes."
RL Proc. Natl. Acad. Sci. U.S.A. 98:4658-4663(2001).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=MGAS315 / Serotype M3;
RX MEDLINE=22133808; PubMed=12122206;
RA Beres S.B., Syva G.L., Barbican K.D., Lei B., Hoff J.S.,
RA Mammarella N.D., Liu M.-Y., Smoot J.C., Porcella S.F., Parkins L.D.,
RA Campbell D.S., Smith T.M., McCormick J.K., Leung D.Y.M.,
RA Schlievert P.M., Musser J.M.;
RT "Genome sequence of a serotype M3 strain of group A Streptococcus:

```

RT	phage-encoded toxins, the high-virulence phenotype, and clone emergence.";
RT	Proc. Natl. Acad. Sci. U.S.A. 99:10078-10083(2002).
RT	[3]
RP	SEQUENCE FROM N.A.
RP	STRAIN=SSI-1 / Serotype M3;
RP	MEDLINE=22683278; PubMed=12799345;
RA	Nakagawa I., Kurokawa K., Yamashita A., Nakata M., Tomiyasu Y.,
RA	Okahashi N., Kawabata S., Yamazaki K., Shiba T., Yasunaga T.,
RA	Hayashi H., Hattori M., Hamada S.;
RT	"Genome sequence of an M3 strain of Streptococcus pyogenes reveals a
RT	large-scale genomic rearrangement in invasive strains and new insights
RT	into phage evolution.";
RL	Genome Res. 13:1042-1055(2003).
RL	[4]
RP	SEQUENCE FROM N.A.
RP	STRAIN=MGAS8232 / Serotype M18;
RP	MEDLINE=21927593; PubMed=11917108;
RA	Smoot J.C., Barbican K.D., Van Gompel J.J., Smoot L.M., Chaussee M.S.,
RA	Sylva G.L., Sturdevant D.E., Ricklefs S.M., Porcella S.F.,
RA	Parkins L.D., Beres S.B., Campbell D.S., Smith T.M., Zhang Q.,
RA	Kapur V., Daly J.A., Veasy L.G., Musser J.M.;
RT	"Genome sequence and comparative microarray analysis of serotype M18
RT	group A Streptococcus strains associated with acute rheumatic fever
RT	outbreaks.";
RL	Proc. Natl. Acad. Sci. U.S.A. 99:4668-4673(2002).
CC	-!- FUNCTION: One of the primary rRNA binding proteins, it binds
CC	directly to 16S rRNA where it nucleates assembly of the body of
CC	the 30S subunit (By similarity).
CC	-!- FUNCTION: With S5 and S12 plays an important role in translational
CC	accuracy (By similarity).
CC	-!- SUBUNIT: Part of the 30S ribosomal subunit. Contacts protein S5.
CC	The interaction surface between S4 and S5 is involved in control
CC	of translational fidelity (By similarity).
CC	-!- SIMILARITY: Contains 1 S4 RNA-binding domain.
CC	-!- SIMILARITY: Belongs to the S4P family of ribosomal proteins.
CC	-----
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC	the European Bioinformatics Institute. There are no restrictions on its
CC	use by non-profit institutions as long as its content is in no way
CC	modified and this statement is not removed. Usage by and for commercial
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC	or send an email to license@isb-sib.ch).
CC	-----
EMBL	AE006635; AAK34809.1; -
EMBL	AE014172; AM80440.1; -
EMBL	AP005146; BAC64924.1; -
EMBL	AE010121; AAL98649.1; -
HSSP	P81288; IC05.
HMAP	MF 01306; -; 1.
InterPro	IPR001912; Ribosomal_S4.
InterPro	IPR005709; Ribosomal_S4_b/o.
InterPro	IPR002942; S4.
Pfam	PF00163; Ribosomal_S4; 1.
Pfam	PF01479; S4; 1.
TIGRFAMs	TIGR01017; rpsD_bact; 1.
PROSITE	PS00632; RIBOSOMAL_S4; 1.
PROSITE	PS00889; S4; 1.
KW	Ribosomal protein; rRNA-binding; Complete proteome.
FT	DOMAIN 93 156 S4 RNA-BINDING.
FT	DOMAIN 93 156 S4 RNA-BINDING.
SEQUENCE	203 AA; 23122 MW; BE4E99E34D7D791D CRC64;
Query Match	26.1%; Score 64; DB 1; Length 203;
Best Local Similarity	31.8%; Pred. No. 0.83;
Matches	14; Conservative 8; Mismatches 22; Indels 0; Gaps 0;
Qy	7 RADELVFLQGLAESREQAKELINAGKVTLTNNSTIPRLKPG 50
Db	94 RLDDVVVRLGLATTRQARQFVNHCHLLVDGKRVDPISYRVDPG 137
RESULT	31

DR HAMAP; MF_01306; -; 1.
 DR InterPro; IPR001912; Ribosomal_S4.
 DR InterPro; IPR005709; Ribosomal_S4_b/o.
 DR InterPro; IPR002942; S4.
 DR Pfam; PF00163; Ribosomal_S4; 1.
 DR Pfam; PF01479; S4; 1.
 DR SMART; SM00363; S4; 1.
 DR TIGRFAMs; TIGR01017; rpsD bact; 1.
 DR PROSITE; PS00632; RIBOSOMAL_S4; 1.
 DR PROSITE; PS00889; S4; 1.
 KW Ribosomal protein; rRNA-binding; Complete proteome.
 FT DOMAIN 96 158 S4 RNA-BINDING.
 SQ SEQUENCE 209 AA; 24084 MW; 2527F85B3BAF09 CRC64;

Query Match 25.7%; Score 63; DB 1; Length 209;
 Best Local Similarity 35.1%; Pred. No. 1.1;
 Matches 13; Conservative 7; Mismatches 17; Indels 0; Gaps 0;

QY 7 RADELVFLQGLAESREQAKRLIMAGKVTLTNNSTTIP 43
 DB 97 RIDNVVVRAGFAISRHAHQIVSHGIIILNGRRVTIP 133

RESULT 34
 RR4_ODOSI STANDARD; PRT; 206 AA.
 AC P49492;
 DT 01-FEB-1996 (Rel. 33, Created)
 DT 01-FEB-1996 (Rel. 33, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Chloroplast 30S ribosomal protein S4.
 GN RPS4.
 OS Odontella sinensis (Marine centric diatom).
 OG Chloroplast.
 OC Eukaryota; stramenopiles; Bacillariophyta; Coscinodiscophyceae;
 OC Biddulphiophycidae; Eupodiscaceae; Odontella.
 OX NCBI_TaxID=2839;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Kowalik K.V., Stoebe B., Schaffran I., Kroth-Pancic P., Freier U.;
 RT "The chloroplast genome of a chlorophyll a+c-containing alga,
 RT Odontella sinensis.";
 RL Plant Mol. Biol. Rep. 13:336-342(1995).
 CC -!- FUNCTION: One of the primary rRNA binding proteins, it binds
 CC directly to 16S rRNA where it nucleates assembly of the body of
 CC the 30S subunit (By similarity).
 CC -!- FUNCTION: With S5 and S12 plays an important role in translational
 CC accuracy (By similarity).
 CC -!- SUBUNIT: Part of the 30S ribosomal subunit. Contacts protein S5.
 CC The interaction surface between S4 and S5 is involved in control
 CC of translational fidelity (By similarity).
 CC -!- SUBCELLULAR LOCATION: Chloroplast.
 CC -!- SIMILARITY: Contains 1 S4 RNA-binding domain.
 CC -!- SIMILARITY: Belongs to the S4P family of ribosomal proteins.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@sb-sib.ch).
 CC
 CC EMBL; Z67753; CAA91654.1; -;
 CC PIR; S78281; S78281.
 CC HSSP; P81288; 1C05.
 CC HAMAP; MF_01306; -; 1.
 DR InterPro; IPR001912; Ribosomal_S4.
 DR InterPro; IPR005709; Ribosomal_S4_b/o.
 DR InterPro; IPR002942; S4.
 DR Pfam; PF00163; Ribosomal_S4; 1.
 DR Pfam; PF01479; S4; 1.
 DR SMART; SM00363; S4; 1.

DR TIGRFAMs; TIGR01017; rpsD bact; 1.
 DR PROSITE; PS00632; RIBOSOMAL_S4; 1.
 DR PROSITE; PS00889; S4; 1.
 KW Ribosomal protein; rRNA-binding; Chloroplast.
 FT DOMAIN 95 157 S4 RNA-BINDING.
 SQ SEQUENCE 206 AA; 23557 MW; F67C0B3FD35969FA CRC64;

Query Match 25.3%; Score 62; DB 1; Length 206;
 Best Local Similarity 30.8%; Pred. No. 1.5;
 Matches 12; Conservative 11; Mismatches 16; Indels 0; Gaps 0;

QY 5 KYRADELVFLQGLAESREQAKRLIMAGKVTLTNNSTTIP 43
 DB 94 EMRLDTICTLGTGAPTASARQLVNEGHITVNDNVVSIP 132

RESULT 35
 RS4_THEAC STANDARD; PRT; 199 AA.
 AC Q9HJD7;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE 30S ribosomal protein S4P.
 GN RPS4P OR TA1032.
 OS Thermoplasma acidophilum.
 OC Archaea; Euryarchaeota; Thermoplasmata; Thermoplasmatales;
 OC Thermoplasmataceae; Thermoplasma.
 OX NCBI_TaxID=2303;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=DSM 1728;
 RX MEDLINE=20479972; PubMed=11029001;
 RA Ruepp A., Graml W., Santos-Martinez M.-L., Koretke K.K., Volker C.,
 RA Mewes H.-W., Frishman D., Stocker S., Lupas A.N., Baumeister W.;
 RT "The genome sequence of the thermoacidophilic scavenger Thermoplasma
 RT acidophilum.";
 RL Nature 407:508-513(2000).
 CC -!- FUNCTION: One of the primary rRNA binding proteins, it binds
 CC directly to 16S rRNA where it nucleates assembly of the body of
 CC the 30S subunit (By similarity).
 CC -!- FUNCTION: With S5 and S12 plays an important role in translational
 CC accuracy (By similarity).
 CC -!- SUBUNIT: Part of the 30S ribosomal subunit. Contacts protein S5.
 CC The interaction surface between S4 and S5 is involved in control
 CC of translational fidelity (By similarity).
 CC -!- SIMILARITY: Contains 1 S4 RNA-binding domain.
 CC -!- SIMILARITY: Belongs to the S4P family of ribosomal proteins.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@sb-sib.ch).
 CC
 CC EMBL; AL445066; CAC12161.1; -;
 CC HAMAP; MF_01306; -; 1.
 DR InterPro; IPR001912; Ribosomal_S4.
 DR InterPro; IPR005710; Ribosomal_S4/9.
 DR InterPro; IPR002942; S4.
 DR Pfam; PF00163; Ribosomal_S4; 1.
 DR Pfam; PF01479; S4; 1.
 DR SMART; SM00363; S4; 1.
 DR TIGRFAMs; TIGR01018; rpsD arch; 1.
 DR PROSITE; PS00632; RIBOSOMAL_S4; 1.
 DR PROSITE; PS00889; S4; 1.
 KW Ribosomal protein; rRNA-binding; Complete proteome.
 FT DOMAIN 106 170 S4 RNA-BINDING.
 SQ SEQUENCE 199 AA; 22927 MW; 1F5C2E0079C5D6F7 CRC64;

Query Match 24.9%; Score 61; DB 1; Length 199;


```
Best Local Similarity 35.1%; Pred. No. 1.9;
Matches 13; Conservative 8; Mismatches 16; Indels 0; Gaps 0;

Qy 7 RADELVFLQGLAESREQAKRLIMAGKVTLTNNSSTIP 43
Db 107 RLQITVFRKNIATSVRQARQITTHGHISVAGRRVTVP 143

RESULT 36
HLVA_TREHY
ID HLVA_TREHY STANDARD; PRT; 240 AA.
AC Q06803;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Hemolysin A.
DE TLVA OR TLY.
OS Treponema hyodysenteriae (Serpulina hyodysenteriae).
OC Bacteria; Spirochaetes; Spirochaetales; Brachyspiraceae; Brachyspira.
OX NCBI_TaxID=159;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=B204;
RX MEDLINE=92112315; PubMed=1730486;
RA Muir S., Koopman M.B.H., Libby S.J., Joens L.A., Heffron F.,
RA Kusters J.G.;
RT "Cloning and expression of a Serpula (Treponema) hyodysenteriae
RL hemolysin gene.";
CC Infect. Immun. 60:529-535 (1992).
CC -!- FUNCTION: Bacterial hemolysins are exotoxins that attack blood
CC cell membranes and cause cell rupture by mechanisms not clearly
CC defined.
CC -!- INDUCTION: By sodium ribonucleate.
CC -!- SIMILARITY: Contains 1 S4 RNA-binding domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X61684; CAA43858.1; -.
CC DR EMBL; A17012; CAA01317.1; -.
CC DR PIR; A43863; A43863.
CC DR InterPro; IPR002877; RtmJ_FtsJ.
CC DR InterPro; IPR002942; S4.
CC DR InterPro; IPR004538; Tly.
CC DR Pfam; PF01728; FtsJ; 1.
CC DR Pfam; PF01479; S4; 1.
CC DR SMART; SM00363; S4; 1.
CC DR TIGRFAMs; TIGR00478; tly; 1.
CC DR TIGRFAMs; TIGR01017; rpsD_bact; 1.
CC DR PROSITE; PS00632; RIBOSOMAL_S4; 1.
CC DR PROSITE; PS50889; S4; 1.
CC DR Ribosomal protein; RNA-binding; rRNA-binding; Chloroplast.
CC FT DOMAIN 1 61 S4 RNA-BINDING.
CC KW Hemolysin; Toxin; RNA-binding.
CC FT DOMAIN 1 61 S4 RNA-BINDING.
CC SQ SEQUENCE 240 AA; 26881 MW; E718A0A21DF3CCB7 CRC64;

Query Match 24.9%; Score 61; DB 1; Length 240;
Best Local Similarity 38.9%; Pred. No. 2.3;
Matches 14; Conservative 7; Mismatches 15; Indels 0; Gaps 0;

Qy 6 VRADELVFLQGLAESREQAKRLIMAGKVTLTNNSST 41
Db 1 MRLDEYHSEGYTSRSKAQIILLACGVFVNGVKVT 36

RESULT 37
RR4_GUTH
ID RR4_GUTH STANDARD; PRT; 200 AA.
AC P17072;
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-AUG-1990 (Rel. 15, Last sequence update)
```

```
28-FEB-2003 (Rel. 41, Last annotation update)
DE Chloroplast 30S ribosomal protein S4.
GN RPS4.
OS Guillardia theta (Cryptomonas phi).
OG Chloroplast.
OC Eukaryota; Cryptophyta; Cryptomonadaceae; Guillardia.
OX NCBI_TaxID=55529;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90245597; PubMed=2336372;
RA Douglas S.E., Durnford D.G.;
RT "Nucleotide sequence of the genes for ribosomal protein S4 and
RT tRNA(Arg) from the chlorophyll c-containing alga Cryptomonas phi.";
RL Nucleic Acids Res. 18:1903-1903 (1990).
CC -!- FUNCTION: One of the primary rRNA binding proteins, it binds
CC directly to 16S rRNA where it nucleates assembly of the body of
CC the 30S subunit (By similarity).
CC -!- FUNCTION: With S5 and S12 plays an important role in translational
CC accuracy (By similarity).
CC -!- SUBUNIT: Part of the 30S ribosomal subunit. Contacts protein S5.
CC The interaction surface between S4 and S5 is involved in control
CC of translational fidelity (By similarity).
CC -!- SUBCELLULAR LOCATION: Chloroplast.
CC -!- SIMILARITY: Contains 1 S4 RNA-binding domain.
CC -!- SIMILARITY: Belongs to the S4p family of ribosomal proteins.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AF041468; AAC35637.1; -.
CC DR HSSP; P81288; 1C05.
CC DR HAMAP; MF_01306; -.
CC DR InterPro; IPR001912; Ribosomal_S4.
CC DR InterPro; IPR005709; Ribosomal_S4_b/o.
CC DR InterPro; IPR002942; S4.
CC DR Pfam; PF00163; Ribosomal_S4; 1.
CC DR Pfam; PF01479; S4; 1.
CC DR SMART; SM00363; S4; 1.
CC DR TIGRFAMs; TIGR01017; rpsD_bact; 1.
CC DR PROSITE; PS00632; RIBOSOMAL_S4; 1.
CC DR PROSITE; PS50889; S4; 1.
CC DR Ribosomal protein; RNA-binding; rRNA-binding; Chloroplast.
CC FT DOMAIN 90 152 S4 RNA-BINDING.
CC SQ SEQUENCE 200 AA; 22846 MW; 3235A3BE16E8B1F2 CRC64;

Query Match 24.5%; Score 60; DB 1; Length 200;
Best Local Similarity 28.2%; Pred. No. 2.6;
Matches 11; Conservative 12; Mismatches 16; Indels 0; Gaps 0;

Qy 5 KVRDELVFLQGLAESREQAKRLIMAGKVTLTNNSSTIP 43
Db 89 EMRLDNVIFRLGMAPTTPAARQLVNHGHKVNTRVSIP 127

RESULT 38
RR4_MARPO
ID RR4_MARPO STANDARD; PRT; 202 AA.
AC P06358;
DT 01-JAN-1988 (Rel. 06, Created)
DT 01-JAN-1988 (Rel. 06, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Chloroplast 30S ribosomal protein S4.
GN RPS4.
OS Marchantia polymorpha (Liverwort).
OG Chloroplast.
OC Eukaryota; Viridiplantae; Embryophyta; Marchantiophyta; Marchantiaceae;
OC Marchantiopsida; Marchantiidae; Marchantiales; Marchantiaceae;
OC Marchantia.
```



```
CC the 30S subunit (By similarity).
CC !- FUNCTION: With S5 and S12 plays an important role in translational
CC accuracy (By similarity).
CC !- SUBUNIT: Part of the 30S ribosomal subunit. Contacts protein S5.
CC The interaction surface between S4 and S5 is involved in control
CC of translational fidelity (By similarity).
CC !- SIMILARITY: Contains 1 S4 RNA-binding domain.
CC !- SIMILARITY: Belongs to the S4P family of ribosomal proteins.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AP005369; BAC07700.1; -
DR HAMAP; MF_01306; -; 1.
DR InterPro; IPR001912; Ribosomal_S4.
DR InterPro; IPR005709; Ribosomal_S4_b/o.
DR InterPro; IPR002942; S4.
DR Pfam; PF00163; Ribosomal_S4; 1.
DR Pfam; PF01479; S4; 1.
DR SMART; SM00363; S4; 1.
DR TIGRFAMs; TIGR01017; rpsD_bact; 1.
DR PROSITE; PS00632; RIBOSOMAL_S4; 1.
DR PROSITE; PS50889; S4; 1.
KW Ribosomal protein; RNA-binding; rRNA-binding; Complete proteome.
FT DOMAIN 90 152 S4 RNA-BINDING.
SQ SEQUENCE 202 AA; 23204 MW; ED9B0049C1C39F09 CRC64;

Query Match 24.5%; Score 60; DB 1; Length 202;
Best Local Similarity 28.3%; Pred. No. 2.6;
Matches 13; Conservative 12; Mismatches 21; Indels 0; Gaps 0;

Qy 5 KVRADLVFLQGLAESREQAKRLMAGKVTLTNNSTTPIRLKPG 50
Db 89 EMRLDNTVFRILGMPTIPAAQLVNHGHLVNGRNVSIPIYQCRPG 134
```

Search completed: July 7, 2004, 16:55:36
Job time : 4.32226 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 7, 2004, 16:51:22 ; Search time 11.6279 Seconds
(without alignments)
1356.728 Million cell updates/sec

Title: US-10-009-919A-1_COPY_1_50

Perfect score: 245

Sequence: 1 MAKVKRADELVLFLGLAES.....GKVTLTNNSTTIPLEKPG 50

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: SPREMBL 25:*

2: sp_archaea:*

3: sp_bacteria:*

4: sp_fungi:*

5: sp_human:*

6: sp_invertebrate:*

7: sp_mammal:*

8: sp_mhc:*

9: sp_organelle:*

10: sp_plant:*

11: sp_rodent:*

12: sp_virus:*

13: sp_vertebrate:*

14: sp_unclassified:*

15: sp_rvirus:*

16: sp_bacteriap:*

17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	97.5	39.8	271	16	Q836w3 enterococcu
2	96	39.2	271	16	Q8dpm1 streptococc
3	93.5	38.2	272	16	Q88wm4 lactobacill
4	93	38.0	275	16	Q99yx6 streptococc
5	93	38.0	275	16	Q8p077 streptococc
6	93	38.0	275	16	Q8k6t0 streptococc
7	93	38.0	279	16	Q8lm55 bacillus an
8	92	37.6	279	16	Q8l8s0 bacillus ce
9	90	36.7	267	16	Q9xlr2 thermotoga
10	90	36.7	275	16	Q8E6l8 streptococc
11	90	36.7	275	16	Q8E163 streptococc
12	89.5	36.5	267	16	Q7tus0 prochloroco
13	88	35.9	274	16	Q92by9 listeria in
14	88	35.9	274	16	Q8y7c0 listeria mo
15	87	35.5	271	16	Q8XJ52 clostridium
16	87	35.5	275	16	Q8dVb2 streptococc

Query Match 39.8%; Score 97.5; DB 16; Length 271;

ALIGNMENTS

RESULT 1

Q836W3 ID Q836W3 PRELIMINARY; PRT; 271 AA.
AC Q836W3;
DT 01-JUN-2003 (TReMBLrel. 24, Created)
DT 01-JUN-2003 (TReMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TReMBLrel. 25, Last annotation update)
DE Hemolysin A.
GN TLYA OR EF0982.
OS Enterococcus faecalis (Streptococcus faecalis).
OC Bacteria; Firmicutes; Lactobacillales; Enterococcaceae; Enterococcus.
OX NCBI_TaxID=1351;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=V583 / ATCC 700802;
RX MEDLINE=22550857; PubMed=12663927;
RA Paulsen I.T., Banerjee L., Myers G.S.A., Nelson K.E., Seshadri R.,
RA Read T.D., Fouts D.E., Eisen J.A., Gill S.R., Heidelberg J.F.,
RA Tettelin H., Dodson R.J., Umayam L., Brinkac L., Beanan M., Nelson W.,
RA Daugherty S., DeBoy R.T., Durkin S., Kolonay J., Madupu R., Khouri H.,
RA Vamathevan J., Tran B., Upton J., Hansen T., Shetty J., Khouri H.,
RA Utterback T., Radune D., Ketchum K.A., Dougherty B.A., Fraser C.M.,
RT "Role of mobile DNA in the evolution of vancomycin-resistant
RT Enterococcus faecalis";
RL Science 299:2071-2074(2003).
DR EMBL; AE016950; AA080788.1; -.
DR TIGR; EF0982; -.
DR GO; GO:0003723; F:RNA binding; IEA.
DR GO; GO:0015070; F:toxin activity; IEA.
DR InterPro; IPR002877; RmJ_FtsJ.
DR InterPro; IPR002942; S4_-.
DR InterPro; IPR004538; Tly.
DR Pfam; PF01728; FtsJ; 1.
DR Pfam; PF01479; S4; 1.
DR TIGRFAMs; TIGR00478; tly; 1.
KW Complete proteome.
SQ SEQUENCE 271 AA; 30062 MW; 777A0AA910F6B511 CRC64;

Q8f969 leptospira
Q8eq47 oceanobacil
Q9k972 bacillus ha
Q97hd6 clostridium
Q7va50 prochloroco
P74319 synechocyst
Q8dix4 synechococc
Q8g5d4 bifidobacte
Q49898 mycobacteri
Q05676 mycobacteri
Q8tya5 methanopyru
Q7ubx7 synechococc
Q8cs54 staphylococ
Q8lkt2 bacillus ce
Q8l7a4 bacillus ce
Q7tu41 prochloroco
Q88ux0 lactobacill
Q9f6y3 chloroflexu
Q8ev12 mycoplasma
Q82z16 enterococcu
Q89rw3 bradyrhizob
Q7vgc0 helicobacte
Q9l3q7 mycobacteri
Q8rac4 thermoaer
Q8a4a1 bacteroides
Q50760 mycobacteri
Q7veu4 mycobacteri
Q8f834 corynebacte
Q66971 aquifex aeo

17 87 35.5 276 16 Q8F969
18 83 33.9 269 16 Q8EQ47
19 81.5 33.3 272 16 Q9K972
20 80 32.7 267 16 Q97HD6
21 77.5 31.6 270 16 Q7VA50
22 77.5 31.6 274 16 P74319
23 76 31.0 262 16 Q8DIX4
24 76 31.0 280 16 Q8G5G4
25 75.5 30.8 268 2 Q49898
26 75.5 30.8 269 16 Q05676
27 74 30.2 79 17 Q8TYA5
28 74 30.2 271 16 Q7UBX7
29 73 29.8 200 16 Q8CS54
30 73 29.8 200 16 Q8LKT2
31 73 29.8 200 16 Q8L7A4
32 72.5 29.6 270 16 Q7TU41
33 72 29.2 202 16 Q88UX0
34 71.5 29.2 260 2 Q9F6Y3
35 71 29.0 202 16 Q8EV12
36 71 29.0 203 16 Q82Z16
37 71 29.0 245 16 Q89RW3
38 70 28.6 208 16 Q7VGC0
39 68.5 28.0 281 2 Q9L3Q7
40 68 27.8 265 16 Q8RAC4
41 67 27.3 201 16 Q8A4A1
42 66.5 27.1 268 16 Q50760
43 66.5 27.1 268 16 Q7VEU4
44 66 26.9 201 16 Q8FS34
45 66 26.9 258 16 Q66971

```
Best Local Similarity 52.2%; Pred. No. 0.0002;
Matches 24; Conservative 4; Mismatches 13; Indels 5; Gaps 1;

QY 5 KVRDELVFQGLAESREQAQRILMAGKVTLTNNSTTIPRLKPG 50
DB 3 KERVDVLAFAQGLFETREKARSMAGLVYNDKNE-----RLDKPG 43

RESULT 2
Q8D9N1 Q8D9N1 PRELIMINARY; PRT; 271 AA.
AC Q8D9N1;
DT 01-MAR-2003 (TReMBLrel. 23, Created)
DT 01-MAR-2003 (TReMBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TReMBLrel. 25, Last annotation update)
DE Conserved hypothetical protein.
GN SPR1086
OS Streptococcus pneumoniae (strain ATCC BAA-255 / R6).
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
NCBI_TaxID=171101;
RN [1]
RP MEDLINE=21429245; PubMed=11544234;
RA Hoskins J., Alborn W.E. Jr., Arnold J., Blaszcak L.C., BURGESS S.,
RA DeHoff B.S., Estrem S.T., Fritz L., Fu D.-J., Fuller W., Geringer C.,
RA Gilmour R., Glass J.S., Khoja H., Kraft A.R., Lagace R.E.,
RA LeBlanc D.J., Lee L.N., Lefkowitz E.J., Lu J., Matsushima P.,
RA McAhren S.M., McHenry M., McLeaster K., Mundy C.W., Niclas T.I.,
RA Norris F.H., O'Gara M., Peery R.B., Robertson G.T., Rockey P.,
RA Sun P.-M., Winkler M.E., Yang Y., Young-Bellido M., Zhao G.,
RA Zook C.A., Baltz R.H., Jaskunas S.R., Rostock P.R. Jr., Skatrud P.L.,
RA Glass J.I.;
RT "Genome of the bacterium Streptococcus pneumoniae strain R6.";
RL J. Bacteriol. 183:5709-5717 (2001).
DR EMBL; AE008481; AAX99889.1; -.
DR PIR; E98007; E98007.
DR GO; GO:0003723; F:RNA binding; IEA.
DR GO; GO:0015070; F:toxin activity; IEA.
DR InterPro; IPR002877; RrmJ_FtsJ.
DR InterPro; IPR002942; S4.
DR InterPro; IPR004538; Tly.
DR Pfam; PF01728; FtsJ; 1.
DR Pfam; PF01479; S4; 1.
DR SMART; SM00363; S4; 1.
DR TIGRFAMs; TIGR00478; tly; 1.
DR PROSITE; PS50889; S4; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 271 AA; 29869 MW; 5EC27C6C43AA99A5 CRC64;

Query Match 39.2%; Score 96; DB 16; Length 271;
Best Local Similarity 50.0%; Pred. No. 0.00032;
Matches 25; Conservative 5; Mismatches 14; Indels 6; Gaps 2;

QY 1 MAKHKVRDELVFQGLAESREQAQRILMAGKVTLTNNSTTIPRLKPG 50
DB 1 MAKERV--DVLAYKQGLFETREQAQRGMAGLVAVLNGE----RFDKPG 44

RESULT 3
Q88WM4 Q88WM4 PRELIMINARY; PRT; 272 AA.
AC Q88WM4;
DT 01-JUN-2003 (TReMBLrel. 24, Created)
DT 01-JUN-2003 (TReMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TReMBLrel. 25, Last annotation update)
DE Hemolysin homolog.
GN LP1603.
OS Lactobacillus plantarum.
OC Bacteria; Firmicutes; Lactobacillales; Lactobacillaceae;
OC Lactobacillus.
NCBI_TaxID=1590;
RN [1]

Best Local Similarity 52.2%; Score 93.5; DB 16; Length 272;
Best Local Similarity 45.7%; Pred. No. 0.00068;
Matches 21; Conservative 8; Mismatches 12; Indels 5; Gaps 1;

QY 5 KVRDELVFQGLAESREQAQRILMAGKVTLTNNSTTIPRLKPG 50
DB 3 KERVDVLLVQOGLFETREKAKRAVMAGELGVNEE-----RLDKPG 43

RESULT 4
Q99YX6 Q99YX6 PRELIMINARY; PRT; 275 AA.
AC Q99YX6;
DT 01-JUN-2001 (TReMBLrel. 17, Created)
DT 01-JUN-2001 (TReMBLrel. 17, Last sequence update)
DT 01-OCT-2003 (TReMBLrel. 25, Last annotation update)
DE Putative hemolysin.
GN HDYAL OR SPY1497.
OS Streptococcus pyogenes.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
NCBI_TaxID=1314;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SF370 / ATCC 700294 / Serotype M1;
RC MEDLINE=21192684; PubMed=11296296;
RA Ferretti J.J., McShan W.M., Ajdic D.J., Savic D.J., Savić G., Lyon K.,
RA Primeaux C., Sezate S., Suvorov A.N., Kenton S., Lai H.S., Lin S.P.,
RA Qian Y., Jia H.G., Najjar F.Z., Ren Q., Zhu H., Song L., White J.,
RA Yuan X., Clifton S.W., Roe B.A., McLaughlin R.;
RT "Complete genome sequence of an M1 strain of Streptococcus pyogenes.";
RC Proc. Natl. Acad. Sci. U.S.A. 98:4658-4663 (2001).
DR EMBL; AE006583; AAK34296.1; -.
DR GO; GO:0003723; F:RNA binding; IEA.
DR GO; GO:0015070; F:toxin activity; IEA.
DR InterPro; IPR002877; RrmJ_FtsJ.
DR InterPro; IPR002942; S4.
DR InterPro; IPR004538; Tly.
DR Pfam; PF01728; FtsJ; 1.
DR Pfam; PF01479; S4; 1.
DR SMART; SM00363; S4; 1.
DR TIGRFAMs; TIGR00478; tly; 1.
DR PROSITE; PS50889; S4; 1.
KW Complete proteome.
SQ SEQUENCE 275 AA; 30440 MW; CA836893EAC91E08 CRC64;

Query Match 38.0%; Score 93; DB 16; Length 275;
Best Local Similarity 50.0%; Pred. No. 0.0006;
Matches 23; Conservative 4; Mismatches 15; Indels 4; Gaps 1;

QY 5 KVRDELVFQGLAESREQAQRILMAGKVTLTNNSTTIPRLKPG 50
DB 3 KERVDVLAFAQGLFETREQAQRGMAGLVVSVINGQ-----RYDKPG 44
```

```
RESULT 5
Q8P077 PRELIMINARY; PRT; 275 AA.
ID Q8P077
AC Q8P077
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Putative hemolysin.
GN SPYM18_1515.
OS Streptococcus pyogenes (serotype M18).
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=186103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MGAS8232 / Serotype M18;
RX MEDLINE=21927593; PubMed=11917108;
RA Smoot J.C., Barbican K.D., Van Gompel J.J., Smoot L.M., Chaussee M.S.,
RA Sylva G.L., Sturdevant D.B., Ricklefs S.M., Porcella S.F.,
RA Parkinson L.D., Beres S.B., Campbell D.S., Smith T.M., Zhang Q.,
RA Kapur V., Daly J.A., Veasy L.G., Mueser J.M.;
RA "Genome sequence and comparative microarray analysis of serotype M18
RT group A Streptococcus strains associated with acute rheumatic fever
RT outbreaks.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:4668-4673 (2002).
DR EMBL; AE010666; AL98086.1; -.
DR GO; GO:0003723; F:RNA binding; IEA.
DR GO; GO:0015070; F:toxin activity; IEA.
DR InterPro; IPR002877; RrmJ_FtsJ.
DR InterPro; IPR002942; S4.
DR InterPro; IPR004538; Tly.
DR Pfam; PF01728; FtsJ; 1.
DR Pfam; PF01479; S4; 1.
DR TIGRFAMs; TIGR00478; tly; 1.
DR PROSITE; PS00889; S4; 1.
KW Complete proteome; Hypothetical protein.
SQ SEQUENCE 275 AA; 30427 MW; 3A350443EAC91E05 CRC64;

Query Match 38.0%; Score 93; DB 16; Length 275;
Best Local Similarity 50.0%; Pred. No. 0.0008;
Matches 23; Conservative 4; Mismatches 15; Indels 4; Gaps 1;

QY 5 KVRADLVFLQGLAESRQAKRLIMAGKVTLTNSTTIPRLKPG 50
| | | | | : | | | | | : | | | | | : | | | | |
Db 3 KERVVDVLAYKQGLFETREQAKRGVMAGLVVSVNGQ----RYDKPG 44

RESULT 6
Q8K6T0 PRELIMINARY; PRT; 275 AA.
ID Q8K6T0
AC Q8K6T0
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Putative hemolysin.
GN HLVA OR SPYM3_1153 OR SPS0709.
OS Streptococcus pyogenes (serotype M3).
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=198466;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MGAS3115 / Serotype M3;
RX MEDLINE=22133808; PubMed=12122206;
RA Beres S.B., Sylva G.L., Barbican K.D., Lei B., Hoff J.S.,
RA Mamavelli N.D., Liu M.-Y., Smoot J.C., Porcella S.F., Parkinson L.D.,
RA Campbell D.S., Smith T.M., McCormick J.K., Leung D.Y.M.,
RA Schlievert P.M., Mueser J.M.;
RA "Genome sequence of a serotype M3 strain of group A Streptococcus:
RT phage-encoded toxins, the high-virulence phenotype, and clone
RT emergence.";
```



```
SQ SEQUENCE 275 AA; 31014 MW; 83D4FBD355247313 CRC64;
Query Match 36.7%; Score 90; DB 16; Length 275;
Best Local Similarity 48.0%; Pred. No. 0.002;
Matches 24; Conservative 6; Mismatches 14; Indels 6; Gaps 2;

Qy 1 MAKHKVRADELVFLQGLAESREQAQLIMAGKVTLTNNSTTIPRLKPG 50
| | | | | : | | | | | : | | | | | : | | | | | : | | | | |
1 MAKERV--DVLAYKQGLFDFTREQAKRGVAGWVINGE---RYDKPG 44

Db

RESULT 11
Q8E163 ID Q8E163 PRELIMINARY; PRT; 275 AA.
AC Q8E163;
DT 01-MAR-2003 (TRENBLrel. 23, Created)
DT 01-MAR-2003 (TRENBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)
DE Hemolysin A.
GN SAG0499.
OS Streptococcus agalactiae (serotype V).
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=216466;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=2603 V/R / Serotype V;
RX MEDLINE=22222988; PubMed=12200547;
RA Tettelin H., Maignani V., Cieslewicz M.J., Eisen J.A., Peterson S.,
RA Wessels M.R., Paulsen I.T., Nelson K.E., Margarit I., Read T.D.,
RA Mañóff L.C., Wolf A.M., Beanan M.J., Brinkac L.M., Daugherty S.C.,
RA DeBoy R.T., Durkin A.S., Kolonay J.F., Madupu R., Lewis M.R.,
RA Radu D., Fedorova N.B., Scanlan D., Khouri H., Mulligan S.,
RA Carty H.A., Cline R.T., Van Aken S.E., Gill J., Scarselli M., Mora M.,
RA Iacobini E.T., Brettoni C., Galli G., Mariani M., Vegni F., Maione D.,
RA Rinaudo D., Rappuoli R., Telford J.L., Kasper D.L., Grandi G.,
RA Fraser C.M.;
RT "Complete genome sequence and comparative genomic analysis of an
RT emerging human pathogen, serotype V Streptococcus agalactiae.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:12391-12396(2002).
DR EMBL; AE014214; AA099401.1; -.
DR TTGR; SAG0499; -.
DR GO; GO:0003723; F:RNA binding; IEA.
DR GO; GO:0015070; F:toxin activity; IEA.
DR InterPro; IPR002877; RrmJ_FtsJ.
DR InterPro; IPR002942; S4.
DR InterPro; IPR004538; Tly.
DR Pfam; PF01728; FtsJ; 1.
DR Pfam; PF01479; S4; 1.
DR SMART; SM00363; S4; 1.
DR TIGRFAMs; TIGR00478; tly; 1.
DR PROSITE; PS50889; S4; 1.
DR Complete proteome.
KW Complete proteome.
SQ SEQUENCE 275 AA; 31014 MW; 83D4FBD355247313 CRC64;

Query Match 36.7%; Score 90; DB 16; Length 275;
Best Local Similarity 48.0%; Pred. No. 0.002;
Matches 24; Conservative 6; Mismatches 14; Indels 6; Gaps 2;

Qy 1 MAKHKVRADELVFLQGLAESREQAQLIMAGKVTLTNNSTTIPRLKPG 50
| | | | | : | | | | | : | | | | | : | | | | | : | | | | |
1 MAKERV--DVLAYKQGLFDFTREQAKRGVAGWVINGE---RYDKPG 44

Db

RESULT 12
Q7TUS0 ID Q7TUS0 PRELIMINARY; PRT; 267 AA.
AC Q7TUS0;
DT 01-OCT-2003 (TRENBLrel. 25, Created)
DT 01-OCT-2003 (TRENBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)
DE FtsJ cell division protein:S4 domain:hemolysin A.
GN PWT1482.
```

```
OS Prochlorococcus marinus (strain MIT 9313).
OC Bacteria; Cyanobacteria; Prochlorophytes; Prochlorococcaceae;
OC Prochlorococcus.
OX NCBI_TaxID=74547;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=22825698; PubMed=12917642;
RA Rocap G., Larimer F.W., Lamerdin J., Malfatti S., Chain P.,
RA Ahlgren N.A., Arellano A., Coleman M., Hauser L., Hess W.R.,
RA Johnson Z.I., Land M., Lindell D., Post A.F., Regala W., Shah M.,
RA Shaw S.L., Steglich C., Sullivan M.B., Ting C.S., Tolonen A.,
RA Webb E.A., Zinser E.R., Chisholm S.W.;
RT "Genome divergence in two Prochlorococcus ecotypes reflects oceanic
RT niche differentiation.";
RL Nature 424:1042-1047(2003).
DR EMBL; BX572099; CAE21657.1; -.
KW Cell division; Complete proteome.
SQ SEQUENCE 267 AA; 29226 MW; 6A261D2762A1335C CRC64;

Query Match 36.5%; Score 89.5; DB 16; Length 267;
Best Local Similarity 46.8%; Pred. No. 0.002;
Matches 22; Conservative 6; Mismatches 14; Indels 5; Gaps 1;

Qy 4 HKVRADELVFLQGLAESREQAQLIMAGKVTLTNNSTTIPRLKPG 50
| | | | | : | | | | | : | | | | | : | | | | | : | | | | |
3 HKQRLDLQLLMKGLASSRHOAQLIRAGKVRDNGQL-----LDKPG 44

Db

RESULT 13
Q92BY9 ID Q92BY9 PRELIMINARY; PRT; 274 AA.
AC Q92BY9;
DT 01-DEC-2001 (TRENBLrel. 19, Created)
DT 01-DEC-2001 (TRENBLrel. 19, Last sequence update)
DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)
DE Hypothetical protein lin1403.
GN Lin1403.
OS Listeria innocua.
OC Bacteria; Firmicutes; Bacillales; Listeriaceae; Listeria.
OX NCBI_TaxID=1642;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CLIP 11262 / Serovar 6a;
RX MEDLINE=21537279; PubMed=11679669;
RA Glaser P., Franceul L., Buchrieser C., Rusniok C., Amend A.,
RA Baquero F., Berche P., Bloecker H., Brandt P., Chakraborty T.,
RA Charbit A., Chetouani F., Couve E., de Daruvar A., Dehoux P.,
RA Domann E., Dominguez-Bernal G., Duchaud E., Durant L., Bussurget O.,
RA Etian K.-D., Feihi H., Garcia-del Portillo F., Garrido P.,
RA Gautier L., Goebel W., Gomez-Lopez N., Hain T., Hauf J., Jackson D.,
RA Jones L.-M., Kaerst U., Krest J., Kuhn M., Kunst F., Kurapkat G.,
RA Madueno E., Maitournam A., Mata Vicente J., Ng E., Nedjari H.,
RA Nordsiek G., Novella S., de Pablo B., Perez-Diaz J.-C., Purcell R.,
RA Remmel B., Rose M., Schluter T., Simoes N., Tierrez A.,
RA Vazquez-Boland J.-A., Voss H., Wehlend J., Cossart P.;
RT "Comparative genomics of Listeria species.";
RL Science 294:849-852(2001).
DR EMBL; AL596168; CAC96634.1; -.
DR FIR; AB1608; AB1608.
DR Listlist; LIN01403; -.
DR GO; GO:0003723; F:RNA binding; IEA.
DR GO; GO:0015070; F:toxin activity; IEA.
DR InterPro; IPR002877; RrmJ_FtsJ.
DR InterPro; IPR002942; S4.
DR InterPro; IPR004538; Tly.
DR Pfam; PF01728; FtsJ; 1.
DR Pfam; PF01479; S4; 1.
DR SMART; SM00363; S4; 1.
DR TIGRFAMs; TIGR00478; tly; 1.
DR PROSITE; PS50889; S4; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 274 AA; 30424 MW; 5FA1FD0E77EFAB9 CRC64;
```

```
Query Match 35.9%; Score 88; DB 16; Length 274;
Best Local Similarity 48.0%; Pred. No. 0.0036;
Matches 24; Conservative 5; Mismatches 15; Indels 6; Gaps 1;

QY 1 MAKHKVRDELVLQGLAESREQAQRLIMAGKVTLTNNSITPIRLLEKPG 50
DB 1 MTIKKERADILLVEQGLFETREKAKRAIMAGIVYRKEE-----RVDKPG 44

RESULT 14
Q8Y7C0 PRELIMINARY; PRT; 274 AA.
AC Q8Y7C0;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein lm01366.
CS Lm01366.
OS Listeria monocytogenes.
OC Bacteria; Firmicutes; Bacillales; Listeriaceae; Listeria.
OX NCBI_TaxID=1639;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=EGD-e / Serovar 1/2a;
RX MEDLINE=21537279; PubMed=11679669;
RA Glaser P., Frangeul L., Blocher H., Brandt P., Chakraborty T.,
RA Quiero F., Berche P., Bucher F., Couve E., de Daruvar A., Dehoux P.,
RA Charbit A., Chetouani F., Duchaud E., Durand L., Dussurget O.,
RA Domann E., Dominguez-Bernal G., Garcia-del Portillo F., Garrido P.,
RA Etian K.-D., Fsihi H., Garcia-del Portillo F., Garrido P.,
RA Gautier L., Goebel W., Gomez-Lopez N., Hain T., Hauf J., Jackson D.,
RA Jones L.-M., Kierst U., Krefz U., Kuhn M., Kunst F., Kurapat G.,
RA Madueno E., Maitoum A., Mata Vicente J., Ng E., Nedjari H.,
RA Nordsiek G., Novella S., de Pablo B., Perez-Diaz J.-C., Purcell R.,
RA Remmel B., Rose M., Schluter T., Simoes N., Tierrez A.,
RA Vazquez-Boland J.-A., Voss H., Wehland J., Cossart P.;
RT "Comparative genomics of Listeria species.";
RL Science 294:849-852(2001).
RC EMBL; AF591978; CAC9444.1; -.
DR PIR; AF1245; AF1245.
DR Listlist; LMO01366; -.
DR GO; GO:0003723; F:RNA binding; IEA.
DR GO; GO:0015070; F:toxin activity; IEA.
DR InterPro; IPR002877; RrmJ_FtsJ.
DR InterPro; IPR002942; S4.
DR InterPro; IPR004538; Tly.
DR Pfam; PF01728; FtsJ; 1.
DR Pfam; PF01479; S4; 1.
DR SMART; SM00363; S4; 1.
DR TIGRFAMs; TIGR00478; tly; 1.
DR PROSITE; PS00889; S4; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 274 AA; 30520 MW; 094A3BC96F567DF1 CRC64;

Query Match 35.9%; Score 88; DB 16; Length 274;
Best Local Similarity 48.0%; Pred. No. 0.0036;
Matches 24; Conservative 5; Mismatches 15; Indels 6; Gaps 1;

QY 1 MAKHKVRDELVLQGLAESREQAQRLIMAGKVTLTNNSITPIRLLEKPG 50
DB 1 MTIKKERADILLVEQGLFETREKAKRAIMAGIVYRKEE-----RVDKPG 44

RESULT 15
Q8XJE2 PRELIMINARY; PRT; 271 AA.
AC Q8XJE2;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Probable hemolysin.
GN HLYD OR CPE1818.
OS Clostridium perfringens.

Query Match 35.9%; Score 88; DB 16; Length 274;
Best Local Similarity 48.0%; Pred. No. 0.0036;
Matches 24; Conservative 5; Mismatches 15; Indels 6; Gaps 1;

QY 1 MAKHKVRDELVLQGLAESREQAQRLIMAGKVTLTNNSITPIRLLEKPG 50
DB 1 MTIKKERADILLVEQGLFETREKAKRAIMAGIVYRKEE-----RVDKPG 44

RESULT 16
Q8DVB2 PRELIMINARY; PRT; 275 AA.
AC Q8DVB2;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Putative hemolysin.
GN SMU.583.
OS Streptococcus mutans.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OX Streptococcus.
OX NCBI_TaxID=1309;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=UAI59 / ATCC 700610 / Serotype C;
RX MEDLINE=22295063; PubMed=12397186;
RA Ajdic D., McShan W.M., McLaughlin R.E., Savic G., Chang J.,
RA Carson M.B., Primeaux C., Tian R., Kenton S., Jia H., Lin S., Qian Y.,
RA Li S., Zhu H., Najjar F., Lal H., White J., Roe B.A., Ferretti J.J.;
RT "Genome sequence of Streptococcus mutans UAI59, a cariogenic dental
RT pathogen.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:14434-14439(2002).
RC EMBL; AR014902; AAN58322.1; -.
DR GO; GO:0003723; F:RNA binding; IEA.
DR GO; GO:0015070; F:toxin activity; IEA.
DR InterPro; IPR002877; RrmJ_FtsJ.
DR InterPro; IPR002942; S4.
DR InterPro; IPR004538; Tly.
DR Pfam; PF01728; FtsJ; 1.
DR Pfam; PF01479; S4; 1.
DR SMART; SM00363; S4; 1.
DR TIGRFAMs; TIGR00478; tly; 1.
DR PROSITE; PS00889; S4; 1.
KW Complete proteome.
SQ SEQUENCE 275 AA; 30546 MW; EEF702E084B88FFD CRC64;

Query Match 35.5%; Score 87; DB 16; Length 275;
Best Local Similarity 47.8%; Pred. No. 0.0048;
Matches 22; Conservative 6; Mismatches 12; Indels 6; Gaps 1;
```



```

[1]
RN  SEQUENCE FROM N.A.
RP  STRAIN=ATCC 824 / DSM 792 / VKM B-1787;
RC  MEDLINE=21359325; PubMed=1146286;
RA  Noelling J., Bregon G., Omelchenko M.V., Makarova K.S., Zeng Q.,
RX  Gibson R., Lee H.M., Dubois J., Qiu D., Hitti J., Wolf Y.I.,
RA  Tatusov R.L., Sabatche F., Doucette-Stamm L., Soucaille P., Daly M.J.,
RA  Bennett G.N., Koonin E.V., Smith D.R.;
RT  "Genome sequence and comparative analysis of the solvent-producing
RT  bacterium Clostridium acetobutylicum.";
RL  J. Bacteriol. 183:4823-4838(2001).
DR  EMBL; AE007710; AAK80035.1; -.
DR  PIR; H97155; H97155.
DR  GO; GO:0008168; F:methyltransferase activity; IEA.
DR  GO; GO:0003723; F:RNA binding; IEA.
DR  GO; GO:0015070; F:toxin activity; IEA.
DR  InterPro; IPR002877; RrmJ_FtsJ.
DR  InterPro; IPR002942; S4.
DR  InterPro; IPR004538; Tly.
DR  Pfam; PF01728; FtsJ; 1.
DR  Pfam; PF01479; S4; 1.
DR  SMART; SM00363; S4; 1.
DR  TIGRFAMs; TIGR00478; tly; 1.
DR  PROSITE; PS50889; S4; 1.
KW  Methyltransferase; Complete proteome.
SQ  SEQUENCE 267 AA; 29924 MW; 7E388BC667C609C3 CRC64;

Query Match 32.7%; Score 80; DB 16; Length 267;
Best Local Similarity 38.0%; Pred. No. 0.039;
Matches 19; Conservative 14; Mismatches 11; Indels 6; Gaps 1;

QY  1 MAKHKVRDELVFLQGLAESREQAKRLIMAGKVTLTNNSTTIPRLKPG 50
Db  1 MSENKERLDVLLVEKGFESREKARASIMAGIYVDD-----LRIDKCG 44

RESULT 21
QY  Q7VA50 PRELIMINARY; PRT; 270 AA.
AC  Q7VA50;
DT  01-OCT-2003 (TrEMBLrel. 25, Created)
DT  01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT  01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE  Predicted rRNA methylase.
GN  PRO1617.
OS  Prochlorococcus marinus.
OC  Bacteria; Cyanobacteria; Prochlorophytes; Prochlorococcaceae;
OC  Prochlorococcus.
OX  NCBI_TaxID=1219;
RN  [1]
RP  SEQUENCE FROM N.A.
RC  STRAIN=SARG / CCMP 1375 / SS120;
RX  MEDLINE=22810154; PubMed=12917486;
RA  Dufresne A., Salanoubat M., Partensky F., Artiguenave F., Axmann I.M.,
RA  Barbe V., Duprat S., Galperin M.Y., Koonin E.V., Le Gall F.,
RA  Makarova K.S., Ostrowski M., Oztas S., Robert C., Rogozin I.B.,
RA  Scanlan D.J., Tandeau de Marsac N., Weissenbach J., Wincker P.,
RA  Wolf Y.I., Hess W.R.;
RT  "Genome sequence of the cyanobacterium prochlorococcus marinus SS120,
RT  a nearly minimal oxyphototrophic genome.";
RL  Proc. Natl. Acad. Sci. U.S.A. 100:10020-10025(2003).
DR  EMBL; AE017165; AAQ00661.1; -.
DR  Methyltransferase; Complete proteome.
SQ  SEQUENCE 270 AA; 29644 MW; 5842FB5EDAA428C4 CRC64;

Query Match 31.6%; Score 77.5; DB 16; Length 270;
Best Local Similarity 45.7%; Pred. No. 0.084;
Matches 21; Conservative 6; Mismatches 14; Indels 5; Gaps 1;

QY  5 KVRDELVFLQGLAESREQAKRLIMAGKVTLTNNSTTIPRLKPG 50
Db  4 KNRLDVHLLTKGLAPSEKQKIRAGKV-----RDVGVNLDKPG 44
```

```

RESULT 22
P74319 PRELIMINARY; PRT; 274 AA.
AC  P74319;
DT  01-FEB-1997 (TrEMBLrel. 02, Created)
DT  01-FEB-1997 (TrEMBLrel. 02, Last sequence update)
DT  01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE  Hypothetical protein slr0950.
GN  slr0950.
OS  Synechocystis sp. (strain PCC 6803).
OC  Bacteria; Cyanobacteria; Chroococcales; Synechocystis.
OX  NCBI_TaxID=1148;
RN  [1]
RP  SEQUENCE FROM N.A.
RX  MEDLINE=97061201; PubMed=8905231;
RA  Kaneko T., Sato S., Kotani H., Tanaka A., Asamizu E., Nakamura Y.,
RA  Miyajima N., Hiroseawa M., Sugiyama M., Sasamoto S., Kimura T.,
RA  Hosouchi T., Matsumoto A., Muraki A., Nakazaki N., Naruo K., Okumura S.,
RA  Shimpou S., Takeuchi C., Wada T., Watanabe A., Yamada M., Yasuda M.,
RA  Tabata S.;
RT  "Sequence analysis of the genome of the unicellular cyanobacterium
RT  Synechocystis sp. strain PCC6803. II. Sequence determination of the
RT  entire genome and assignment of potential protein-coding regions.";
RL  DNA Res. 3:109-136(1996).
DR  EMBL; D90914; BAA18413.1; -.
DR  PIR; S76154; S76154.
DR  GO; GO:0003723; F:RNA binding; IEA.
DR  GO; GO:0015070; F:toxin activity; IEA.
DR  InterPro; IPR002877; RrmJ_FtsJ.
DR  InterPro; IPR002942; S4.
DR  InterPro; IPR004538; Tly.
DR  Pfam; PF01728; FtsJ; 1.
DR  Pfam; PF01479; S4; 1.
DR  SMART; SM00363; S4; 1.
DR  TIGRFAMs; TIGR00478; tly; 1.
DR  PROSITE; PS50889; S4; 1.
KW  Hypothetical protein; Complete proteome.
SQ  SEQUENCE 274 AA; 30026 MW; ADD3AF2B4E425A89 CRC64;

Query Match 31.6%; Score 77.5; DB 16; Length 274;
Best Local Similarity 43.1%; Pred. No. 0.085;
Matches 22; Conservative 7; Mismatches 15; Indels 7; Gaps 2;

QY  1 MAK-HKVRDELVFLQGLAESREQAKRLIMAGKVTLTNNSTTIPRLKPG 50
Db  1 MAKADKQRLDALLVAGKLCESRALAQRLIRAGEVKVNNQ-----LVDKPG 45

RESULT 23
Q8DIX4 PRELIMINARY; PRT; 262 AA.
AC  Q8DIX4;
DT  01-MAR-2003 (TrEMBLrel. 23, Created)
DT  01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT  01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE  Hemolysin-like protein.
GN  TLL1457.
OS  Synechococcus elongatus (Thermosynechococcus elongatus).
OC  Bacteria; Cyanobacteria; Chroococcales; Synechococcus.
OX  NCBI_TaxID=32046;
RN  [1]
RP  SEQUENCE FROM N.A.
RX  STRAIN=BP-1;
RX  MEDLINE=22225144; PubMed=12240834;
RA  Nakamura Y., Kaneko T., Sato S., Ikeuchi M., Katoh H., Sasamoto S.,
RA  Watanabe A., Iriguchi M., Kawashima K., Kimura T., Kishida Y.,
RA  Kiyokawa C., Kohara M., Matsumoto M., Matsumoto A., Nakazaki N.,
RA  Shimpou S., Sugimoto M., Takeuchi C., Yamada M., Tabata S.;
RT  "Complete genome structure of the thermophilic cyanobacterium
RT  Thermosynechococcus elongatus BP-1.";
RL  DNA Res. 9:123-130(2002).
DR  EMBL; AP005374; BAC09009.1; -.

```

```
DR GO; GO:0003723; F:RNA binding; IEA.
DR GO; GO:0015070; F:toxin activity; IEA.
DR InterPro; IPR002877; RrmJ_FtsJ.
DR InterPro; IPR002942; S4.
DR InterPro; IPR004538; Tly.
DR Pfam; PF01728; FtsJ; 1.
DR SMART; SM00363; S4; 1.
DR TIGRFAMs; TIGR00478; tly; 1.
DR PROSITE; PS0889; S4; 1.
KW Complete proteome.
SQ SEQUENCE 262 AA; 28607 MW; 1B5291F0B133A27B CRC64;

Query Match 31.0%; Score 76; DB 16; Length 262;
Best Local Similarity 42.0%; Pred. No. 0.13;
Matches 21; Conservative 8; Mismatches 15; Indels 2; Gaps 2;

QY 1 MAKHKVRADLVFLOGLAESREQAKRLIMAGKVTLTNNSTTIPRLKPKG 50
DB 1 MTPRKQRLDSLVERHLCESRQQAQRWIRAGAVHVNH----IP--IDKPKG 44

RESULT 24
Q8G5G4 PRELIMINARY; PRT; 280 AA.
AC Q8G5G4;
DT 01-MAR-2003 (TRENBLrel. 23, Created)
DT 01-MAR-2003 (TRENBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)
DE Hemolysin-like protein with S4 domain found in bacteria and
DE plants.
DE GN BL1048.
OS Bifidobacterium longum.
OC Bacteria; Actinobacteria; Actinobacteridae; Bifidobacteriales;
OC Bifidobacteriaceae; Bifidobacterium.
OX NCBI_TaxID=216816;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NCC 2705;
RX MEDLINE=22294977; PubMed=12381787;
RA Schell M.A., Karmirantzou M., Snel B., Vilanova D., Berger B.,
RA Pessi G., Zwahlen M.-C., Desiere F., Bork P., Delley M.,
RA Pridmore R.D., Arigoni F.;
RT "The genome sequence of Bifidobacterium longum reflects its adaptation
RT to the human gastrointestinal tract.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:14422-14427(2002).
DR EMBL; AE014728; AAN24854.1; -.
DR GO; GO:0003723; F:RNA binding; IEA.
DR GO; GO:0015070; F:toxin activity; IEA.
DR InterPro; IPR002877; RrmJ_FtsJ.
DR InterPro; IPR002942; S4.
DR InterPro; IPR004538; Tly.
DR Pfam; PF01728; FtsJ; 1.
DR SMART; SM00363; S4; 1.
DR TIGRFAMs; TIGR00478; tly; 1.
DR PROSITE; PS0889; S4; 1.
KW Complete proteome.
SQ SEQUENCE 280 AA; 29817 MW; 7DA44D7D5BA1EEB4 CRC64;

Query Match 31.0%; Score 76; DB 16; Length 280;
Best Local Similarity 40.9%; Pred. No. 0.14;
Matches 18; Conservative 7; Mismatches 19; Indels 0; Gaps 0;

QY 7 RADELVFLQGLAESREQAKRLIMAGKVTLTNNSTTIPRLKPKG 50
DB 37 RLDMMLVASGLVESRAKARLIKAGHVRVDGETITKPSFMVKAG 80

RESULT 25
Q49898 PRELIMINARY; PRT; 268 AA.
ID Q49898
AC Q49898;
DT 01-NOV-1996 (TRENBLrel. 01, Created)
DT 01-NOV-1996 (TRENBLrel. 01, Last sequence update)
DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)
DE U0247a.
OS Mycobacterium leprae.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacteriaceae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1769;
RN [1]
RP SEQUENCE FROM N.A.
RA Smith D.R.;
RL Submitted (JAN-1994) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Robison K.;
RL Submitted (SEP-1994) to the EMBL/GenBank/DBJ databases.
DR EMBL; U00021; A0450906.1; -.
DR PIR; S72968; S72968.
DR GO; GO:0003723; F:RNA binding; IEA.
DR GO; GO:0015070; F:toxin activity; IEA.
DR InterPro; IPR002877; RrmJ_FtsJ.
DR InterPro; IPR002942; S4.
DR InterPro; IPR004538; Tly.
DR Pfam; PF01728; FtsJ; 1.
DR SMART; SM00363; S4; 1.
DR TIGRFAMs; TIGR00478; tly; 1.
DR PROSITE; PS0889; S4; 1.
SQ SEQUENCE 268 AA; 28063 MW; 099EE1BD6F667AC9 CRC64;

Query Match 30.8%; Score 75.5; DB 2; Length 268;
Best Local Similarity 51.4%; Pred. No. 0.15;
Matches 18; Conservative 8; Mismatches 8; Indels 1; Gaps 1;

QY 1 MAKHKVRADLVFLOGLAESREQAKRLIMAGKVTL 35
DB 1 MAR-RVRVDVLRRLRSRQAQKLISAGKVS 34

RESULT 26
O05676 PRELIMINARY; PRT; 269 AA.
ID O05676
AC O05676;
DT 01-JUL-1997 (TRENBLrel. 04, Created)
DT 01-JUL-1997 (TRENBLrel. 04, Last sequence update)
DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)
DE Hypothetical protein (Cytotoxin/hemolysin).
DR GN TLXA OR ML1358 OR MLC1351.14C.
OS Mycobacterium leprae.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacteriaceae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1769;
RN [1]
RP SEQUENCE FROM N.A.
RA Murphy L., Harris D.;
RL Submitted (MAY-1997) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=TN;
RX MEDLINE=21128732; PubMed=11234002;
RA Cole S.T., Eiglmeier K., Parkhill J., James K.D., Thomson N.R.,
RA Wheeler P.R., Honore N., Garnier T., Churcher C., Harris D.,
RA Mungall K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R.M., Devlin K., Duthoy S., Felwell T., Fraser A., Hamlin N.,
RA Holroyd S., Hornsby T., Jagels K., Lacroix C., Maclean J., Moule S.,
RA Murphy L., Oliver K., Quail M.A., Rajadream M.A., Rutherford K.M.,
RA Rutter S., Seeger K., Simon S., Simmonds M., Skelton J., Squares R.,
RA Squares S., Stevens K., Taylor K., Whitehead S., Woodward J.R.,
RA Barrell B.G.;
RT "Massive gene decay in the leprosy bacillus.";
RL Nature 409:1007-1011(2001).
DR EMBL; Z95117; CAB08287.1; -.
DR EMBL; AL503921; CAC31739.1; -.

```

```
DR PIR; H87078; H87078.
DR Leproma; ML1358; -.
DR GO; GO:0003723; F:RNA binding; IEA.
DR GO; GO:0015070; F:toxin activity; IEA.
DR InterPro; IPR002877; RtmJ_FtsJ.
DR InterPro; IPR002942; S4.
DR InterPro; IPR004538; Tly.
DR Pfam; PF01728; FtsJ; 1.
DR SMART; SM00363; S4; 1.
DR TIGRFAMs; TIGR00478; tly; 1.
DR PROSITE; PS50889; S4; 1.
DR Hypothetical protein; Complete proteome.
KW SEQUENCE 269 AA; 28203 MW; 738B78D8AF5EA043 CRC64;

Query Match 30.8%; Score 75.5; DB 16; Length 269;
Best Local Similarity 51.4%; Pred. No. 0.15;
Matches 18; Conservative 8; Mismatches 1; Gaps 1;

QY 1 MAKHKVRADELVFQGLAESREQAKRLIMAGKVTL 35
|||:|||||:|||||:|||||:|||||:
Db 1 MAR-RVRVDVLRGLARSRQQAKLISAGKVS 34

RESULT 27
Q8TYAS PRELIMINARY; PRT; 79 AA.
AC Q8TYAS;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Predicted RNA-binding protein containing the S4 domain.
GN MK0398.
OS Methanopyrus kandleri.
OC Archaea; Euryarchaeota; Methanopyri; Methanopyrales; Methanopyraceae;
OC Methanopyrus
OX NCBI_TaxID=2320;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=AV19 / DSM 6324 / JCM 9639;
RX MEDLINE=21927647; PubMed=11930014;
RA Slesarev A.I., Mezhevaya K.V., Makarova K.S., Polushin N.N.,
RA Shcherbinina O.V., Shakhova V.V., Belova G.I., Aravind L.,
RA Natale D.A., Rogozin I.B., Tatusov R.L., Wolf Y.I., Stetter K.O.,
RA Malykh A.G., Koonin E.V., Kozaykin S.A.;
RT "The complete genome of hyperthermophile Methanopyrus kandleri AV19
RT and monophyly of archaeal methanogens."
RL Proc. Natl. Acad. Sci. U.S.A. 99:4644-4649(2002).
DR EMBL; AE010336; AA01613.1; -.
DR GO; GO:0003723; F:RNA binding; IEA.
DR InterPro; IPR002942; S4.
DR Pfam; PF01479; S4; 1.
DR SMART; SM00363; S4; 1.
DR PROSITE; PS50889; S4; 1.
KW Complete proteome.
SQ SEQUENCE 79 AA; 8930 MW; 87B143AA16BAF785 CRC64;

Query Match 30.2%; Score 74; DB 17; Length 79;
Best Local Similarity 40.9%; Pred. No. 0.057;
Matches 18; Conservative 6; Mismatches 20; Indels 0; Gaps 0;

QY 7 RADELVFQGLAESREQAKRLIMAGKVTLTNNSTTIPRLRKPG 50
|||:|||||:|||||:|||||:|||||:
Db 6 RLDAFLDVGLAESRRAKLIVESGRVNGKLVRKPLVSPG 49

RESULT 28
Q7U8X7 PRELIMINARY; PRT; 271 AA.
AC Q7U8X7;
DT 01-OCT-2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
```

```
DE Hemolysin-like protein.
GN SYNW0482.
OS Synchococcus sp. (strain WH8102).
OC Bacteria; Cyanobacteria; Chroococcales; Synchococcus.
OX NCBI_TaxID=84588;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=22825697; PubMed=12917641;
RA Palenik B., Brahamsha B., Larimer F.W., Land M., Hauser L., Chain P.,
RA Lamerdin J., Regala W., Allen E.B., McCarren J., Paulsen I.,
RA Dufresne A., Partensky F., Webb E.A., Waterbury J.;
RT "The genome of a motile marine Synchococcus."
RL Nature 424:1037-1042(2003).
DR EMBL; BX569690; CAB06997.1; -.
KW Complete proteome.
SQ SEQUENCE 271 AA; 29030 MW; CD3BC1E6B46EBB4A CRC64;

Query Match 30.2%; Score 74; DB 16; Length 271;
Best Local Similarity 40.0%; Pred. No. 0.24;
Matches 22; Conservative 9; Mismatches 14; Indels 10; Gaps 2;

QY 5 KVRADLVFLQGLAESREQAKRLIMAGKV-----TL-----TNNSTTIPRLRKPG 49
|||:|||||:|||||:|||||:|||||:
Db 4 KQRLDELRLSRLVRSRQQAQLIRACKVRDGAAGTLLDKPCTEVAALRLRVEQP 58

RESULT 29
Q8CS54 PRELIMINARY; PRT; 200 AA.
AC Q8CS54;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE 30S ribosomal protein S4.
GN SBL396.
OS Staphylococcus epidermidis.
OC Bacteria; Firmicutes; Bacillales; Staphylococcus.
OX NCBI_TaxID=1282;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 12228;
RA Zhang Y., Ren S., Li H., Fu G., Lu L., Lu G., Jia J., Tu Y., Qin Z.,
RA Chen Z., Wen Y.;
RL Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AE016748; AA004995.1; -.
DR GO; GO:0005622; C:intracellular; IEA.
DR GO; GO:0015935; C:small ribosomal subunit; IEA.
DR GO; GO:0003723; F:RNA binding; IEA.
DR GO; GO:0003735; F:structural constituent of ribosome; IEA.
DR InterPro; IPR001912; Ribosomal_S4.
DR InterPro; IPR005709; Ribosomal_S4_b/o.
DR InterPro; IPR002942; S4.
DR Pfam; PF00163; Ribosomal_S4; 1.
DR Pfam; PF01479; S4; 1.
DR SMART; SM00363; S4; 1.
DR TIGRFAMs; TIGR01017; rpsD_bact; 1.
DR PROSITE; PS50889; S4; 1.
KW Ribosomal protein; Complete proteome.
SQ SEQUENCE 200 AA; 23105 MW; E63F28209D66A776 CRC64;

Query Match 29.8%; Score 73; DB 16; Length 200;
Best Local Similarity 36.4%; Pred. No. 0.23;
Matches 16; Conservative 8; Mismatches 20; Indels 0; Gaps 0;

QY 7 RADELVFQGLAESREQAKRLIMAGKVTLTNNSTTIPRLRKPG 50
|||:|||||:|||||:|||||:|||||:
Db 93 RLDAVYSLGLARTTRQARQLVNHGIEVDGGRVDIPSYSLKPG 136

RESULT 30
Q81KT2 PRELIMINARY; PRT; 200 AA.
```


DT 01-JUN-2003 (TReMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TReMBLrel. 25, Last annotation update)
DE Ribosomal protein S4.
GN RPSD OR IP 2331.
OS Lactobacillus plantarum.
OC Bacteria; Firmicutes; Lactobacillales; Lactobacillaceae;
OC Lactobacillus.
OC Lactobacillus.
OC NCBI_TaxID=1590;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NCIMB 8826 / WCFS1;
RX MEDLINE=22480296; PubMed=12566566;
RA Kleerebezem M., Boekhorst J., van Kranenburg R., Molenaar D.,
RA Kuipers O.P., Leer R., Tarchini R., Peters S.A., Sandbrink H.M.,
RA Fiers M.W.E.J., Stekema W., Klein Lankhorst R.M., Bron P.A.,
RA Hoffer S.M., Nierop Groot M.N., Kerkhoven R., De Vries M., Ursing B.,
RA De Vos W.M., Slezep R.J.;
RT "Complete genome sequence of Lactobacillus plantarum WCFS1.";
RL Proc. Natl. Acad. Sci. U.S.A. 100:1990-1995(2003).
DR ENBL; AL935258; CAD64654.1; -.
DR GO: GO:0005622; C:intracellular; IEA.
DR GO: GO:0005840; C:ribosome; IEA.
DR GO: GO:0003735; F:RNA binding; IEA.
DR GO: GO:0003735; F:structural constituent of ribosome; IEA.
DR GO: GO:0006412; P:protein biosynthesis; IEA.
DR InterPro: IPR001912; Ribosomal_S4.
DR InterPro: IPR002942; S4.
DR Pfam: PF00163; Ribosomal_S4; 1.
DR Pfam: PF01479; S4; 1.
DR PROSITE; PS00632; RIBOSOMAL_S4; 1.
DR PROSITE; PS00889; S4; 1.
DR Complete proteome.
KW PROSITE; PS00889; S4; 1.
SQ SEQUENCE 202 AA; 22963 MW; E9E35010EAFSA7 CRC64;

Query Match 29.4%; Score 72; DB 16; Length 202;
Best Local Similarity 34.7%; Pred. No. 0.31; Mismatches 16; Indels 8; Gaps 1;
Matches 17; Conservative 8;

QY 3 KHKV-----RADELVFLQGLAESREQAQLIMAGKVLTTNNSTTIP 43
DB 82 KHGVNFWMLERLDNNVYRLGLATTRQARQLVNHGHITVDGKRVIP 130

RESULT 34
QY9F6Y3 PRELIMINARY; PRT; 260 AA.
AC Q9F6Y3;
DT 01-MAR-2001 (TReMBLrel. 16, Created)
DT 01-MAR-2001 (TReMBLrel. 16, Last sequence update)
DT 01-OCT-2003 (TReMBLrel. 25, Last annotation update)
DE HlyA.
GN HLYA.
OS Chloroflexus aurantiacus.
OC Bacteria; Chloroflexi; Chloroflexales; Chloroflexaceae; Chloroflexus.
OC NCBI_TaxID=1108;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20433268; PubMed=10976061;
RA Xiong J., Fischer W.M., Inoue K., Nakahara M., Bauer C.E.;
RT "Molecular evidence for the early evolution of photosynthesis.";
RL Science 289.1724-1730(2000).
DR ENBL; AF288458; AG15202.1; -.
DR GO: GO:0003723; F:RNA binding; IEA.
DR GO: GO:0015070; F:toxin activity; IEA.
DR InterPro: IPR002877; RrmJ_FtsJ.
DR InterPro: IPR002942; S4.
DR InterPro: IPR004538; Tly.
DR Pfam: PF01728; FtsJ; 1.
DR Pfam: PF01479; S4; 1.
DR SMART; SM00363; S4; 1.
DR TIGRFAMs; TIGR00478; tly; 1.
DR PROSITE; PS00889; S4; 1.
SQ SEQUENCE 260 AA; 27667 MW; 730C8B47AA3E5D99 CRC64;

Query Match 29.2%; Score 71.5; DB 2; Length 260;
Best Local Similarity 33.3%; Pred. No. 0.49;
Matches 18; Conservative 12; Mismatches 13; Indels 11; Gaps 1;

QY 7 RADELVFLQGLAESRQAQLIMAGKV-----TLTNNSTTIPRLLEKP 49
DB 5 RLDQVLVSRGLAETRAQAQLIMAGQVLVNGQVQTKAGTLIADDAVEVRTGLP 58

RESULT 35
QY8EV12 PRELIMINARY; PRT; 202 AA.
AC Q8EV12;
DT 01-MAR-2003 (TReMBLrel. 23, Created)
DT 01-MAR-2003 (TReMBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TReMBLrel. 25, Last annotation update)
DE Ribosomal protein S4.
GN MYPEV550.
OS Mycoplasma penetrans.
OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.
OC NCBI_TaxID=28227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=HF-2;
RX MEDLINE=22354719; PubMed=12466555;
RA Sasaki Y., Ishikawa J., Yamashita A., Oshima K., Kenri T., Furuya K.,
RA Yoshino C., Horino A., Shiba T., Sasaki T., Hattori M.;
RT "The complete genomic sequence of Mycoplasma penetrans, an
RT intracellular bacterial pathogen in humans.";
RL Nucleic Acids Res. 30:5293-5300(2002).
DR ENBL; AP004173; BAC44549.1; -.
DR GO: GO:0005622; C:intracellular; IEA.
DR GO: GO:0015935; C:small ribosomal subunit; IEA.
DR GO: GO:0003723; F:RNA binding; IEA.
DR GO: GO:0003735; F:structural constituent of ribosome; IEA.
DR GO: GO:0006412; P:protein biosynthesis; IEA.
DR InterPro: IPR001912; Ribosomal_S4.
DR InterPro: IPR005709; Ribosomal_S4_b/o.
DR InterPro: IPR002942; S4.
DR Pfam: PF00163; Ribosomal_S4; 1.
DR Pfam: PF01479; S4; 1.
DR SMART; SM00363; S4; 1.
DR TIGRFAMs; TIGR01017; rpsD_bact; 1.
DR PROSITE; PS00632; RIBOSOMAL_S4; 1.
DR PROSITE; PS00889; S4; 1.
KW Complete proteome.
SQ SEQUENCE 202 AA; 23219 MW; 6483477E20B75775 CRC64;

Query Match 29.0%; Score 71; DB 16; Length 202;
Best Local Similarity 29.7%; Pred. No. 0.42;
Matches 19; Conservative 11; Mismatches 18; Indels 16; Gaps 2;

QY 1 MAXH-----KYRADELVFLQGLAESREQAQLIMAGKVLTTNNSTTIP---L 44
DB 76 IAKHMKGALTLNTFIALESRLDNLVYRMGFAPTRRAARQLVNHGHILLDGKVTIPSCMV 135

QY 45 RLEK 48
DB 136 KLEQ 139

RESULT 36
QY82Z16 PRELIMINARY; PRT; 203 AA.
AC Q82Z16;
DT 01-JUN-2003 (TReMBLrel. 24, Created)
DT 01-JUN-2003 (TReMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TReMBLrel. 25, Last annotation update)
DE Ribosomal protein S4.
GN RPSD OR EF3070.
OS Enterococcus faecalis (Streptococcus faecalis).
OC Bacteria; Firmicutes; Lactobacillales; Enterococcaceae; Enterococcus.

OX NCBI_TaxID=1351;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=V583 / ATCC 700802;
RX MEDLINE=22550857; PubMed=12663927;
RA Paulsen I.T., Banerjee L., Myers G.S.A., Nelson K.E., Seshadri R.,
RA Read T.D., Fouts D.E., Eisen J.A., Gill S.R., Heidelberg J.F.,
RA Tettelin H., Dodson R.J., Unayak L., Brinkac L., Beanan M.,
RA Vamatheswari S., DeBoy R.T., Durkin S., Kolonay J., Madupu R., Nelson W.,
RA Dathatharan J., Tran B., Upton J., Hansen T., Shetty J., Khouri H.,
RA Utterback T., Radune D., Ketchum K.A., Dougherty B.A., Fraser C.M.;
RT "Role of mobile DNA in the evolution of vancomycin-resistant
RT Enterococcus faecalis.";
RL Science 299:2071-2074 (2003).
DR EMBL; AB016956; AAC82752.1; --
DR TIGR; EF3070; --
DR GO; GO:0005622; C:intracellular; IEA.
DR GO; GO:0015935; C:small ribosomal subunit; IEA.
DR GO; GO:0003723; F:RNA binding; IEA.
DR GO; GO:0003735; F:structural constituent of ribosome; IEA.
DR GO; GO:0006412; P:protein biosynthesis; IEA.
DR InterPro; IPR001912; Ribosomal_S4.
DR InterPro; IPR005709; Ribosomal_S4_b/o.
DR InterPro; IPR002942; S4.
DR Pfam; PF00163; Ribosomal_S4; 1.
DR Pfam; PF01479; S4; 1.
DR SMART; SM00363; S4; 1.
DR TIGRFAMs; TIGR01017; rpsd bact; 1.
DR PROSITE; PS00632; RIBOSOMAL_S4; 1.
DR PROSITE; PS0889; S4; 1.
KW Ribosomal protein; Complete proteome.
SQ SEQUENCE 203 AA; 2323 MW; E372215A9E5E0093 CRC64;

Query Match 29.0%; Score 71; DB 16; Length 203;
Best Local Similarity 34.7%; Pred. No. 0.43; Mismatches 8; Indels 1;
Matches 17; Conservative 8;

QY 3 KHKV-----RADELVFLQGLAESREQAKRLIMAGKVTLTNNSTTIP 43
DB 62 KHGVNFMVLLEQLNDVNVYRLGLATRRQARQLVNHGHITVDGKRVDP 130

RESULT 37
Q89RW3
ID Q89RW3 PRELIMINARY; PRT; 245 AA.
AC Q89RW3
DT 01-JUN-2003 (TREMBLrel. 24, Created)
DT 01-JUN-2003 (TREMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE B112649 protein.
GN B112649.
OS Bradyrhizobium japonicum.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Bradyrhizobiaceae; Bradyrhizobium.
OX NCBI_TaxID=375;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=USDA 110;
RX MEDLINE=22484998; PubMed=12597275;
RA Kaneko T., Nakamura Y., Sato S., Minamisawa K., Uchiumi T.,
RA Sasamoto S., Watanabe A., Idegawa K., Iriguchi M., Kawashima K.,
RA Kohara M., Matsumoto M., Shimpo S., Tsuruoka H., Wada T., Yamada M.,
RA Tabata S.;
RT "Complete genomic sequence of nitrogen-fixing symbiotic bacterium
RT Bradyrhizobium japonicum USDA110.";
RL DNA Res. 9:189-197 (2002).
DR EMBL; AF005944; BAC47914.1; --
DR GO; GO:0003723; F:RNA binding; IEA.
DR InterPro; IPR002877; RmJ_FtsJ.
DR InterPro; IPR002942; S4.
DR Pfam; PF01728; FtsJ; 1.
DR Pfam; PF01479; S4; 1.
DR PROSITE; PS50889; S4; 1.

KW Complete proteome.
SQ SEQUENCE 245 AA; 26173 MW; EC3915FE562C75B5 CRC64;

Query Match 29.0%; Score 71; DB 16; Length 245;
Best Local Similarity 41.9%; Pred. No. 0.53; Mismatches 18; Indels 0; Gaps 0;
Matches 18; Conservative 7;

QY 1 MAKHKVRADELVFLQGLAESREQAKRLIMAGKVTLTNNSTTIP 43
DB 2 MSPSRKRAILLVERGLFESRARARAIEAGLVTTADDDKQVTKP 44

RESULT 38
Q7VGC0
ID Q7VGC0 PRELIMINARY; PRT; 208 AA.
AC Q7VGC0
DT 01-OCT-2003 (TREMBLrel. 25, Created)
DT 01-OCT-2003 (TREMBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE Ribosomal protein S4.
GN RPSD OR H1402.
OS Helicobacter hepaticus.
OC Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacteriales;
OC Helicobacteriaceae; Helicobacter.
OX NCBI_TaxID=32025;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 51449 / 3B1;
RX MEDLINE=22709201; PubMed=12810954;
RA Suerbaum S., Josenhans C., Sterzenbach T., Drescher B., Brandt P.,
RA Bell M., Droge M., Partmann B., Fischer H.-P., Ge Z., Hoerster A.,
RA Holland R., Klein K., Koenig J., Macko L., Mendz G.L., Nyakatura G.,
RA Schauer D.B., Shen Z., Weber J., Frosch M., Fox J.G.;
RT "The complete genome sequence of the carcinogenic bacterium
RT Helicobacter hepaticus.";
RL Proc. Natl. Acad. Sci. U.S.A. 100:7901-7906 (2003).
DR EMBL; AB017148; AAP77999.1; --
KW Ribosomal protein; Complete proteome.
SQ SEQUENCE 208 AA; 24051 MW; 0EBF61F4E061DE46 CRC64;

Query Match 28.6%; Score 70; DB 16; Length 208;
Best Local Similarity 31.8%; Pred. No. 0.59; Mismatches 14; Conservative 9; Indels 0; Gaps 0;
Matches 14; Conservative 9;

QY 7 RADELVFLQGLAESREQAKRLIMAGKVTLTNNSTTIPRLRLEKPG 50
DB 99 RLNDVVRMGFATRRFARQLVTHGHILVNGKRVDPISYMWKPG 142

RESULT 39
Q9L307
ID Q9L307 PRELIMINARY; PRT; 281 AA.
AC Q9L307
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE Putative haemolysin.
GN TLYA.
OS Mycobacterium ulcerans.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacteriaceae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1809;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 19423;
RA Mve Obiang A., Gomez Lopez A., Portals F., Fonteyne P.A.;
RT "Cloning and expression of a Mycobacterium ulcerans hemolysin gene.";
RL Submitted (JAN-2000) to the EMBL/GenBank/DBSJ databases.
DR EMBL; AJ271681; CAB83047.1; --
DR GO; GO:0003723; F:RNA binding; IEA.
DR GO; GO:0015070; F:toxin activity; IEA.
DR InterPro; IPR002877; RmJ_FtsJ.
DR InterPro; IPR002942; S4.

Search completed: July 7, 2004, 16:56:46
Job time : 14.6279 secs